DOCUMENT RESUME

ED 173 586 | **CE 021 980**

AUTHOR	Borus, Michael E.
TITLE	Measuring the Impact of Employment-Related Social Programs. A Primer on the Evaluation of Employment and Training, Vocational Education, Vocational Rehabilitation, and Other Job-Oriented Programs.
INSTITUTION	Upjohn (W.E.) Inst. for Employment Research, Kalamazoo, Mich.
PUB DATE	Apr 79
NOTE	148p.
AVAILABLE FROM	W. E. Upjohn Institute for Employment Research, 300 South Westnedge Ave., Kalamazoo, Michigan 49007 ($3.50; 10-25 copies, 10% off; 26-50, 15%; 51-100, 20%; over 100, 25%)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	Cost Effectiveness; *Employment Programs; *Evaluation Methods; Program Costs; Program Effectiveness; *Program Evaluation; Program Planning; Research Design; *Social Services; Vocational Education; Vocational Rehabilitation
IDENTIFIERS	*Program Impact

ABSTRACT

An approach and methodology for the systematic measurement of the impact of employment-related social programs is presented in this primer. Chapter 1 focuses on evaluation as the third step (the first two being planning and operation) in the process of program implementation. Chapter 2 examines the impacts of social programs. Topics include defining the goals and impacts of employment-related programs, source of impact definitions, defining whose impacts will be measured, and the listing of potential benefits from employment-related programs. Chapter 3 discusses evaluation design. Presented is information on different designs, the choice of an experimental group, the timing of impact measurement, collecting data on impacts, sample size, and the choice of independent variables. Chapter 4 examines the costs of social programs. Topics include estimating opportunity-costs, measuring the increment in costs, calculating marginal costs, and examining the costs of the first program participants. Finally, chapter 5 presents a suggested schema for comparing program benefits and costs. Discussion includes assigning value to future impacts and costs, accounting for externalities, and making program decisions. Concluding the chapter is a summary outline for evaluating the impact of employment-related programs. (JH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED173586

MEASURING THE IMPACT OF EMPLOYMENT-RELATED SOCIAL PROGRAMS

A Primer on the Evaluation of Employment and Training, Vocational Education, Vocational Rehabilitation, and Other Job-Oriented Programs

Michael E. Borus

"PERMISSION TO REPRODUCE THIS
MATERIAL IN MICROFICHE ONLY
HAS BEEN GRANTED BY
Upjohn Institute
for Employment Research
TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

April 1979

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

THE W. E. UPJOHN INSTITUTE FOR EMPLOYMENT RESEARCH

CE 021 980

Library of Congress Cataloging in Publication Data

Borus, Michael E
Measuring the impact of employment-related social programs.

First ed. published in 1970 under title: Measuring the impact of manpower programs.
Includes bibliographical references.
1. Manpower policy—United States—Evaluation.
I. Title.
HD5724.B625 1979 331.1'1'0973 79-13151

Copyright © 1979
by the
W.E. UPJOHN INSTITUTE
FOR EMPLOYMENT RESEARCH

300 South Westnedge Ave.
Kalamazoo, Michigan 49007

THE INSTITUTE, a nonprofit research organization, was established on July 1, 1945. It is an activity of the W. E. Upjohn Unemployment Trustee Corporation, which was formed in 1932 to administer a fund set aside by the late Dr. W. E. Upjohn for the purpose of carrying on "research into the causes and effects of unemployment and measures for the alleviation of unemployment."

This study is for sale by the Institute at $3.50 per copy. For quantity orders of this publication or any combination of Institute publications, price reductions are as follows: 10-25 copies, 10 percent; 26-50, 15; 51-100, 20; over 100, 25.

THE BOARD OF TRUSTEES OF THE W.E. UPJOHN UNEMPLOYMENT TRUSTEE CORPORATION

Preston S. Parish, Chairman
Mrs. Ray T. Parfet, Vice Chairman
Charles C. Gibbons, Vice Chairman
D. Gordon Knapp, Secretary-Treasurer
E. Gifford Upjohn, M.D.
Donald S. Gilmore
Mrs. Genevieve U. Gilmore
James H. Duncan
John T. Bernhard

THE STAFF OF THE INSTITUTE

E. Earl Wright, Director
Saul J. Blaustein
Judith K. Brawer
Phyllis R. Buskirk
H. Allan Hunt
Paul J. Kozlowski
John R. Mekemson
Philip M. Scherer
Robert A. Straits
Jack R. Woods

FOREWORD

Timely and accurate evaluations of social programs are essential to sound planning and public policy decisions at local, state, and national levels. The need is clearly evident in the field of employment-related programs, which involve the expenditure of billions of dollars and affect the lives of millions of individuals.

This primer presents an approach and methodology for the systematic measurement of the impact of employment-related social programs. The primary emphasis is on basic techniques of evaluation, with references to numerous theoretical and conceptual issues. This guide should add significantly to the literature on program evaluation and is intended to assist those who conduct impact evaluations as well as program planners and administrators who must make decisions based on such evaluations.

Facts and observations as presented in this monograph are the sole responsibility of the author. His viewpoints do not necessarily represent positions of the W. E. Upjohn Institute for Employment Research.

E. Earl Wright
Director

Kalamazoo, Michigan
April 1979

THE AUTHOR

Michael E. Borus, director of the Center for Human Resource Research and professor of Labor and Human Resources at The Ohio State University, received his Ph.D. from Yale University. He has been deputy director of the Office of Research and Development, Employment and Training Administration, U.S. Department of Labor, and has served as consultant to the Office of the Assistant Secretary for Policy, Evaluation and Research, the Office of Youth Programs, and the U.S. Employment Service, U.S. Department of Labor. Borus is the author of *The Economic Effectiveness of Retraining the Unemployed,* co-author of *The Economic Benefits and Costs of Retraining,* and editor of *Evaluating the Impact of Manpower Programs,* as well as numerous articles in the field of program evaluation.

PREFACE

This book is a revised version of an earlier work, *Measuring the Impact of Manpower Programs: A Primer,* written by the present author and William R. Tash and published by the Institute of Labor and Industrial Relations, The University of Michigan-Wayne State University as Policy Papers in Human Resources and Industrial Relations No. 17. The present volume has been changed considerably from the earlier primer. These changes are based on the development of the field in the ten years since the first work was written.

Many persons contributed to this volume. The author would like especially to thank Bill Tash for his earlier contributions and his comments on this version; the many students in classes conducted at Michigan State University, the University of Utah and The Ohio State University whose questions and reactions helped to clarify the presentation; and the many reviewers of the drafts of this and the previous version—Paul Barton, Thomas Bruening, C. Gregory Buntz, John Cheston, Steven Director, Ronald Ehrenberg, Eli Ginzberg, Audrey Freedman, Patricia Greene, Frank Lewis, Garth Mangum, Herbert Parnes, Edward Prescott, Gosta Rehn, Blossom Reutlinger, Harold Sheppard, Frank Shuler, Abraham Stahler, Ernst Stromsdorfer, Ralph Walker, Barbara Weinstein, and Alfred Zuck.

This book is dedicated to the memory of three men who introduced me to the evaluation of social programs: E. Wight Bakke, Joseph Borus, and Gerald G. Somers.

Michael E. Borus

Columbus, Ohio
April 1979

CONTENTS

Foreword iv

Preface vi

Chapter 1 Introduction 1

What is Evaluation? 1
Why Evaluate? 2
What Are the Types of Evaluation? 3
Who Evaluates? 6
Organization of This Primer 7
Chapter Outline 8
Exercise 1-1 11

Chapter 2 The Impacts of Social Programs 13

Defining the Goals and Impacts of Employment-Related Programs 13
Final Versus Intermediate Outcomes as Impacts 14
Sources of Impact Definitions 15
Defining Whose Impacts Will Be Measured 18
Listing of Potential Benefits From Employment-Related Programs 20
Discussion of Criteria 27
The Choice of Impacts and Criteria 32
Exercise 2-1 34
Exercise 2-2 34

Chapter 3 Evaluation Design 35

Alternative Designs 35
  The Case Study Design 36
  Before-and-After Design 37
  Comparison Group Designs 39

  - Random Control Groups 42
  - Practical Problems of Random Assignment 43
  - Theoretical Problems with Random Assignment 46
  - Alternatives to Random Assignment 50
- The Choice of an Experimental Group 52
- The Timing of Impact Measurement 54
- The Choice of Retrospective or Concurrent Data Collection 55
- Collecting Data on the Impacts 56
  - Social Security Administration Data 57
  - Tax Data 59
  - Unemployment Insurance Wage Data 59
- Sample Size 60
- Sources of Data on Comparison Group Members 61
  - Current Population Survey 62
  - Social Security Administration CWHS Data 63
  - National Longitudinal Surveys 64
- The Choice of Independent Variables for Analysis 66
  - Personal Independent Variables 66
  - Program Component Independent Variables 70
  - Exogenous Independent Variables 70
- Determining Proper Program Size—Measuring at the Margin 70
- Exercise 3-1 73
- Exercise 3-2 74

Chapter 4 The Costs of Social Programs 79

- Estimating Opportunity Costs 79
- Listing of Costs of Employment-Related Programs 82
- Measuring the Increment in Costs 87
  - The Use of Control Groups 88
  - Measuring Costs for Other Family Members 89
- Accurately Measuring the Costs 90
  - Proper Program Assignment 90
  - Inclusion of All Costs 91
- Calculating Marginal Costs 92
- Joint Costs 93

Examining the Costs of the First Program Participants 94
Exercise 4-1 96
Exercise 4-2 100

Chapter 5 Combining the Measures of Program Impact and Cost 101

Treating Future Impacts 101
  How the Impacts Might Occur Over Time 102
  The Duration of Impacts 104
  A Possible Solution 104
Assigning Values to Future Impacts and Costs 105
Accounting for Externalities 109
  Displacement Effects 111
  Vacuum Effects 113
  Substitution Effects 114
  Complementary Effects 115
  Measuring the Externalities 115
  Secondary Impacts 116
Making Program Decisions 117
  Combinations of Program Impact and Cost Measures .. 118
  Should a Program be Continued? 121
  Comparison of Alternative Programs 121
  Improving Program Efficiency 123
  Matching Clients and Programs 124
Steps for Evaluating the Impact of Employment-Related Programs: A Summary 124
Limitations on Evaluation Techniques 129
A Final Word for Impact Evaluation 131
Exercise 5-1 132
Exercise 5-2 133

References 135

Chapter 1
INTRODUCTION

What is Evaluation?

Evaluation is the systematic gathering of information in order to make choices among alternative courses of action. In terms of social programs, evaluation can be looked upon as the third step in the process of program implementation (see Chart 1-1). The first step is planning. Planning consists of the definition of the social problems to be attacked and the choice of the course of action to be taken in the solution of the problems. The second step is program operations. This is an attempt to solve the problem through the program outlined in the plan. The third step is evaluation which seeks to determine whether the program was successful in solving the problem or could be more successful than it was.

Chart 1-1. The Process of Program Implementation

Planning	Program operation	Evaluation
Definition of a social problem and the development of a program to solve the problem	An attempt to solve the social problem through a specific course of action	Measuring whether the program successfully solved the social problem

11

Obviously, there are links between evaluation and the other two steps in the process. Planning must be based on information gathered in past evaluations. Through the study of the success of past programs we develop better future means to solve our problems because we need not repeat errors made by others, and we can expand those aspects of past programs which have proved to be successful. Similarly, program operations can be altered to maximize the successful elements of the program and to remove or change those facets which are unsuccessful. In this way, evaluation permits us constantly to improve and upgrade the quality and mix of social programs by providing feedback on successes and failures.

Why Evaluate?

One reason for conducting evaluations has just been presented. Man, as a thinking, rational animal, will seek to choose among alternative courses of action so as to maximize his well being. By systematically gathering information on the success and failure of his past actions, he improves his ability to get the most that he can (maximizes his benefits) while giving up as little as necessary (minimizes his costs) in his present and future actions.

In terms of social programs, however, there is usually another reason for conducting evaluations—there are external pressures or requirements for such evaluations. Such pressures arise from the relative newness and controversiality of social programs. These programs are large in cost, in number of people affected, and in their perceived abilities to do good or harm. Consequently, evaluation of social programs is increasingly required by federal and state legislation.[1] These same factors also have forced program operators to justify the continuation or expansion of their program both to their superiors and to the public through analytical results. Finally, various innovations in public sector budgeting, particularly the program-planning-budgeting system (PPBS) and zero-based budgeting (ZBB), require a greater analytical justification than was necessary under earlier budgeting systems.

---

[1] Examples of legislative requirements for evaluation may be found in Section 313 and Section 316 of the Comprehensive Employment and Training Act, as amended; Section 401 of the Rehabilitation Act of 1973, and Section 112 of the Vocational Education Act of 1963, as amended. For an overview on federal evaluation of social programs, see Wholey, *et al.* (36).

What Are the Types of Evaluation?

The evaluation of social programs can be divided into three general types, each of which asks different questions but can be thought of as a continuum of steps to trace out the programs' effects. 1) *Process evaluation* asks the question "How did the program operate?" 2) *Impact evaluation* asks the question "What did the program do?" 3) *Strategic analysis* seeks to answer the question "How effective was this program in solving the social problem as compared with all of the other programs directed at the problem?"[2]

Process evaluation compares the manner in which a program is operated and the products it produces against the plan for the program. It is plan oriented—it tests whether the plan is being carried out as written, on the basis that the plan must be followed in order to have success.[3] It is operations oriented—it is concerned with what goes on in the program and what the program does to and for the participants.

Process evaluation begins with the assumption that in order to solve a problem, certain preconditions must occur.[4] Therefore, one type of process evaluation (referred to as *control* evaluation) is to test for the existence of these factors. For example, state educational authorities usually require a bachelors degree and specific training for teachers to be certified to teach in the public schools. School districts are then examined to insure that their teachers are certified. There is in this evaluation the implicit assumption that in order for youth to learn, their teachers must meet the state requirements. Another example would be the requirement that the participants in manpower programs be unemployed, underemployed, or economically disadvantaged. The assumption is made that in order for individuals in these categories to benefit from training and other employment-related programs, they must be the persons who receive the services. The control type of process evaluation examines whether the inputs to the program meet a set of predetermined standards.

---

[2] This is just one of many taxonomies for describing the types of evaluation. Others may be found in Scriven (28), whose *formative evaluation* is similar to process evaluation and *summative evaluation* roughly corresponds to our impact evaluation. Ferman (15) also has an excellent categorization.

[3] In the case of program failure, we want to know "Was the idea (the plan) wrong or was it the execution of the idea?" Process evaluation seeks to determine whether the execution was correct, which implies that the plan was inappropriate, or whether the execution was faulty, so that the idea was not truly tested.

[4] These preconditions may be specified in the plan (or regulations) or in commonly accepted management theory. They usually look at aspects of the question, "*Who* should do *what, when,* for *whom* and *with how much?*"

A second type of process evaluation is *monitoring*. The emphasis here is on meeting predetermined standards; however, monitoring focuses on accomplishments or outputs rather than inputs. Monitoring would ask such questions as "Has a youngster increased his reading level by a full year during the time he was in the third grade?" or "Have the participants in employment and training programs found jobs and increased their incomes?"

Impact evaluation seeks to measure the effects of the program. It tries to answer the question "What difference has the program made?" The emphasis is on the changes brought about by the existence of the program. Impact evaluation seeks to compare what occurs, given the existence of the social program, to what would have occurred if the program did not exist. For example, impact evaluation seeks to determine how much better a child reads after a year of remedial tutoring in reading as compared to what his reading level would have been if he had been left in his regular classroom for that period. It is important to note that we are measuring the change caused by the remedial tutoring. We wish to know not only how much progress occurred but also whether this was more than would have taken place had there been no tutoring. In the case of employment-related programs, we would want to know if the income of a participant was higher during or after the program than it would have been had he not entered the program.

It should also be noted that impact evaluation seeks to measure all of the differences caused by the program and is not limited to only those which were originally listed as goals for the program in its plan. It is quite possible that social programs will have outcomes other than those originally planned. For instance, it has been found that while increasing education leads, as expected, to more productive populations and more satisfied and capable citizenry, it also is associated with lower fertility rates.

Finally, strategic analysis seeks to compare alternative programs in order to judge their relative efficiency at accomplishing long-run, large-scale outcomes. Such analysis might compare the ultimate effect on the productivity of the society of an increase in expenditures for primary education and manpower training. Strategic analysis compares the results of impact evaluations for more than one program.

In addition to the type of question being asked, another feature which distinguishes the types of evaluation is the time period for which the program is evaluated. Control evaluation examines the structure of the program and its inputs; consequently, it looks at information which is gathered during the period of program operation. Monitoring also takes place during the operation of the program, but it usually will also gather information about post-program results. Impact evaluation requires a longer period of time after the conclusion of a program in order to determine what the long-run results are. Strategic analysis compares the results of impact evaluations for a variety of competing programs and must wait until all of the impact evaluations have been conducted.

The distinctions between the various types of evaluation may be clearer if we use an example from the medical field:

1. A child is born. We check that the nursery is clean and well staffed by trained personnel, and that the child is fed on a regular schedule.

   This is control evaluation. It examines inputs to the process and measures them against standards of necessity.

2. As the child grows older, we periodically measure its height and weight. We also see if the child walks by the time he or she is two years old.

   This is monitoring. It examines progress and output against predetermined standards of what should occur.

3. The child is innoculated against various diseases. We wish to measure whether the innoculation has led to a longer and more productive life.

   This is impact evaluation. It relates changes in outputs to a change in the inputs.

4. An evaluator attempts to determine if an expenditure on public health supervision, primary medical care, or medical research is the most efficient way of extending the longevity of the population.

   This is strategic analysis. It compares the efficiency of alternative strategies for accomplishing a goal.

Who Evaluates?

Types of evaluation may also be differentiated by the kinds of agencies that conduct them and the uses to which they are put. Since control evaluations basically deal with whether prescribed conditions exist or do not exist, it is the easiest type of evaluation to conduct and does not require an intimate knowledge of the program's operations. Thus, it can be conducted by individuals outside of the program, usually by accountants and auditors. This type of evaluation is often associated with the General Accounting Office (GAO), the "Federal Representative" who reviews local programs, and other "outsiders" who usually attempt to assess programs quickly. On the other hand, monitoring usually involves the program operators or members of the program's administrative staff since one of the basic purposes of monitoring is to provide the program operator with early feedback on his successes and failures. Monitoring questions are often built into the program's management information system.

Impact evaluation is usually done by an external agency. Most program operators do not have the skills, budget, time, or interest necessary to conduct impact evaluations. Further, internal evaluations may give the appearance of conflict of interest. Hence, contracts for such evaluations are often let by federal agencies to private researchers or universities.

Strategic analysis can be conducted only at the highest levels since it compares programs that cut across normal agency and jurisdictional lines. For instance, in the federal executive branch, the Office of Management and Budget (OMB) performs this function. In Congress, the House and Senate Budget Committees are responsible for allocating funds across broad subject areas, while the Appropriations Committees divide the funds among competing programs. The Congressional Budget Office (CBO) is charged with providing the analyses on which these decisions are made.[5]

Organization of This Primer

This primer is concerned with impact evaluation; process evaluation is already covered well in other works,[6] and we shall not cover strategic analysis since our purpose is to introduce the reader to the basics of evaluation techniques. In addition, the primer concentrates on questions dealing with the mechanics of conducting evaluations and omits most of the theoretical discussions which have occurred in the field of evaluation (although references to these discussions are provided). While the techniques for evaluating social programs are not difficult, most past evaluations have not included all of those basic components necessary to arrive at reliable policy decisions. By giving special emphasis to the areas in which past studies have failed, we hope to present an elementary discussion which will help correct the apparent deficiencies in those past studies.

---

[5] The types of evaluation are obviously not mutually exclusive. As noted earlier, in the case of a failure (determined by an impact evaluation) we want to know whether the theory of the plan or its execution was faulty (which requires process evaluation). If we have a success (again using an impact evaluation), we will want to make sure that the success was due to the plan being followed and not to some deviation from the plan.

[6] See in particular *Focus on Manpower Planning* (16), Menzi (21), and U.S. Department of Labor (34).

Readers can best use this primer if they attempt to test their knowledge and make use of it as they go along. To encourage this approach, exercises are included at the end of each chapter. We also suggest that readers examine several of the studies cited in each section. Each of these was selected because it presented theoretical arguments in greater detail than was possible in the space available. Hopefully, this primer, together with the exercises and references, will serve as a jumping-off point in the evaluative process, permitting evaluators to develop their own analyses without committing the same mistakes that have marred earlier studies.

Chapter Outline

The purpose of the evaluation process is to provide policy makers with the basic data necessary for them to make decisions wisely. Impact evaluations of social programs examine the long-run outcomes and view success and failure in these terms. They should provide five essential sets of information. First, they should provide the data necessary to determine if a particular program should be continued. Second, they should determine which of alternative programs achieve the greatest gains for a given cost. Third, evaluations should present information on the components of each program and the mixes of components which are most effective for a given expenditure so that maximum operating efficiency can be achieved. Fourth, evaluations should provide the first three types of information for participants with different characteristics so that a decision maker may determine which individuals are best served by each program. Finally, in the course of evaluating existing programs, data should be gathered which will suggest new methods for attacking social problems. Few impact evaluations of social programs have provided all of this information to date.

One of the major problems in the evaluation of employment-related social programs is that these programs encompass a wide variety of desired outcomes for the nation's workers and potential workers. Generally, they seek to improve the employment situation of the program clients, and in this way to better their economic, physical, and mental well-being. The programs also seek to increase the productive ability of the nation's human resources and to reduce poverty and social dependency. These goals, however, are broad and difficult to operationalize. As a consequence, evaluations of employment-related programs often have been narrow in focus, usually limited to the most obvious effects of the programs such as the average increment in earnings of participants or the number of participants placed in training-related jobs. Other less apparent but possibly important impacts have frequently been ignored. Therefore, our first task is to define more of the basic objectives and impacts of employment-related programs and attempt to establish criteria to measure these.

Other problems which have arisen in social impact program evaluations revolve around the question "Whom do employment-related programs affect?" It appears that past studies have excluded, often because of a lack of data, many persons whose labor market experience was influenced by such programs. In Chapter 2 we point out some of the other groups that should also be examined.

The designs for measuring the success of employment-related programs in past studies have often been weak. Many of the problems have arisen in the use of control or comparison groups. To estimate the effects of a program, it is necessary to compare the experience of the program participants with that of some reference group whose experience can be said to represent what would have happened to the participants in the absence of the program. Unfortunately, results of past studies which were contrary to the prejudices of policy makers have too often been dismissed on the grounds of noncomparability between program participants and "controls." In Chapter 3, we present a procedure designed to solve the comparability problems. We also supply descriptive information on possible sources of data which might be useful for measuring the benefits of employment-related programs.

The problems involved in measuring the costs of social programs have been similar to those involved in measuring program success. Past studies often have not measured all of the appropriate costs, have inadequately selected control groups for cost analysis, and have ignored some groups who incur costs. These are the issues discussed in Chapter 4.

A final problem which has limited the usefulness of many previous evaluations has been the lack of comparability in the presentation of the results of these studies. Chapter 5 presents a suggested schema for comparing program benefits and costs. We also suggest in that chapter a number of technical aspects which have often been lacking in previous studies—the use of multivariate analysis to separate the influences of the wide variety of possible determinants of program success, the use of marginal analysis whenever possible, and the methods for projecting and discounting future effects of the programs. Finally, we present a summary outline of the procedural steps discussed in earlier chapters.

Exercise 1-1

Label each of the following types of evaluation questions as process (control or monitoring), impact, or strategic analysis.

1. Do all vocational rehabilitation counselors have degrees in counseling?
2. Does each student welder have 15 square feet of floor space?
3. Are vocational education program graduates placed in jobs which make use of the skills taught to them in school?
4. Are no more than 10 percent of registrants in a "holding status" at any point in time?
5. Has a public service job increased the income of the person hired?
6. Does vocational rehabilitation lead to happier, more satisfied clients?
7. Is training or work experience a better way to increase employability of youth?

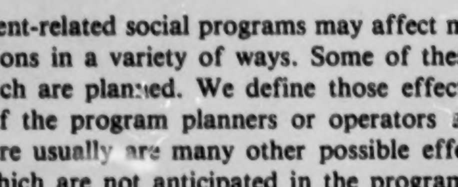
Chapter 2

THE IMPACTS OF SOCIAL PROGRAMS

Defining the Goals and Impacts of Employment-Related Programs

Employment-related social programs may affect many persons and institutions in a variety of ways. Some of these are direct benefits which are planned. We define those effects which are objectives of the program planners or operators as "program goals." There usually are many other possible effects of these programs which are not anticipated in the program plan either because they are side effects or occur to persons who are not directly involved in the program. Some of these are positive outcomes; others may do harm. These unanticipated effects, together with the program goals, we call "impacts." Thus, as shown in Chart 2-1, program goals are a subset of the program's impacts.

Chart 2-1
Impacts of Employment-Related Programs



In our opinion, an attempt should be made to measure all of the impacts and not just the goals. Rationale for this opinion appears below.

Final Versus Intermediate Outcomes as Impacts

In considering impacts, it is useful to distinguish between those which are ultimate objectives and those which are only intermediate steps to achieving those objectives. For instance, higher annual earnings for participants may be the ultimate goal for a program; intermediate outcomes which are sometimes examined to measure this goal are increased wage rates at placement and increased number of persons placed in training related jobs. The intermediate outcomes can be differentiated in two respects: they occur sooner and/or they are only partial measures of the impact.

The need for a time distinction is obvious. It is usually necessitated by an inability to wait until all of the consequences of program participation have occurred. (This is natural when programs can affect the entire lifetimes of participants.) There is a danger, which will be discussed further in chapter 5, that early results may not be indicative of longer term consequences. Shortly after the program, the participants may find themselves at a disadvantage if they have been taken out of the labor market for a substantial period. On the other hand, they may have gained a short-run advantage by making special contacts during the program.

More seriously, intermediate outcomes may not measure the same factors as the ultimate objective. For instance, the absence of higher wage rates may not mean annual earnings are not higher since the program may increase hours of employment. Likewise, persons may be placed in the jobs for which they were trained but these jobs may pay less than other jobs which might have been held. In these cases the implicit assumptions that the intermediate outcomes were identical or highly correlated with the ultimate objective are not correct.

For these reasons it is useful to define impacts in terms of ultimate objectives whenever possible. Similarly, care should be taken when establishing measurement criteria to use those which most closely reflect the ultimate objectives. The further removed what is being measured is from the eventual outcome, in concept and in time, the greater will be the chance of inadequate or improper estimates of the true impacts.

Sources of Impact Definitions

The first order of business in conducting an impact evaluation is to define the broad impacts and the more specific criteria which may be used to judge the effectiveness of the programs. Where do we find these impacts and criteria?

The obvious place to begin is with the goals as defined in the legislation of established programs to see what the drafters thought the program would accomplish. Unfortunately, this very often leads to statements that are difficult to operationalize with criteria. For instance, the Comprehensive Employment and Training Act of 1973 (CETA) had the following Statement of Purpose:

> It is the purpose of this Act to provide job training and employment opportunities for economically disadvantaged, unemployed, and underemployed persons, and to assure that training and other services lead to maximum employment opportunities and enhance self-sufficiency by establishing a flexible and decentralized system of Federal, State, and local programs.

This statement describes the organization of the program (decentralized), the type of participants (unemployed, underemployed, and economically disadvantaged persons), and the types of services to be performed (job training and employment), but is vague as to the outcomes to be achieved (maximum employment opportunities and enhanced self-sufficiency). A review of a few other pieces of legislation will indicate that this lack of clearly stated legislative goals is not the exception.[1]

Other sources of information on potential impacts are the legislative hearings held prior to the establishment of the program and the hearings held on appropriations. Again, however, one is usually left with vague statements of overall objectives. Moreover, the goals of a program may change over time from those originally stated in the legislative process. For instance, the Manpower Development and Training Act of 1962 was originally passed to combat displacement caused by automation, but several years later it was used primarily as a program for integrating minority groups into the labor force without changing the legislation.

One can also turn to the operators of the program, program clients, and other potential users of the evaluation to obtain their perceptions of the program's goals. It is important to note, however, that the goals of program clients are likely to be highly individualized and limited to their personal desires. Likewise, program managers often think more about the services which they must provide and the techniques for providing these services than about the impacts of providing these services. Therefore, the suggestions both groups offer are often incomplete and narrow in their viewpoint.

The agency funding the evaluation also may offer recommendations on which potential impacts to consider. A problem with relying on this source to define the objectives to be studied is the possibility of being co-opted. Scriven (29) argues for "goal-free evaluation" where the evaluator selects the criteria for evaluating a program completely independently of the program staff.[2] Such independence, however, may allow the evaluator to overlook some goals and criteria. Knowing the goals of the staff need not necessarily bias the evaluator.

---

1. For an excellent statement of why this occurs, see Donnelly (14). It should be noted that Congress recognized the vagueness of their purpose when they amended CETA in 1978. There they added a phrase saying CETA should result in an increase in the earned income of the economically disadvantaged, unemployed, or underemployed.

2. Also see his interview in Salasin (27).

Since these sources are limited in their ability to supply the evaluator with all of the impacts to be examined, it is incumbent on him to think of all the possible areas in which the program could conceivably have an effect. That many of these will not be stated objectives for the program is not important. For example, there is nothing in the CETA legislation which addresses the program's effects on worker health. Yet, CETA may have an impact in this area because: 1) program participants have increased contact with social service agencies; 2) part of the additional incomes may be spent on health services if the program is successful in raising worker incomes; and 3) a substantial portion of the training under the Act will be in the area of health services. Thus, while health was not mentioned by the authors of the Act, it might be very much affected by CETA. The impact in this area may be considered very important by some people using the program's evaluations for decision making.

As a general rule, it is much better to attempt to measure impacts which prove not to exist than to ignore impacts which do exist. There are two reasons for this. First, both the political and economic scene may change as time passes, and the goals of the program may change as was seen with the Manpower Development and Training Act of 1962. If the new goals were not included in the evaluation, it may now be worthless since it will not answer the questions being asked under the new conditions. Second, since models of the labor market are very incomplete, they are unable to predict with certainty where the impact of various employment-related programs will be felt. Social scientists are unable to model all of the relationships or to state what the effect of a particular action will be. As more and more evaluations are conducted and as a wider variety of possible impacts is examined, the models will undoubtedly improve.

Thus, while it is valuable to determine who will be using the evaluation and what its uses are to be, getting complete insights on the success of the programs under evaluation usually makes it imperative to go beyond these considerations and to include all of the possible impacts. Great care should be taken before an objective is eliminated from consideration due to the particular uses for which the evaluation is purported to be made. A change in the political party or the economic situation may drastically change the value structure used to rank the alternative impacts. The evaluator must remember that, while it is always possible to disregard information once collected, it will require an entirely new evaluation to test for the effects of an impact which has been omitted from the study.

Defining Whose Impacts Will Be Measured

In establishing the benefits to be studied it is extremely important to note that programs affect different groups in different ways and at different levels. We can identify four primary parties who may benefit from employment-related programs. These are society as a whole, participants in the program, employers, and the government. Each of the groups has different goals which they wish the program to accomplish. Therefore, depending on the perspective taken, the objectives of the program will differ. From a societal point of view, the goals of employment-related programs are put in terms of aggregate changes. Examples of societal goals would be increased aggregate production, improved equity in the overall distribution of income and employment, and reductions in the national unemployment rate. For the individual participant the goals are more limited, usually to benefits which directly affect him, such as improvement in his earnings and increases in his satisfaction. An employer, too, will tend to look at the programs in terms of his own interests. For instance, he will be concerned with how the productivity of his labor force has been increased. Finally, the government will view the programs in terms of the various societal objectives but, in addition, will seek programs which aid its budgetary position by increasing revenues and reducing expenditures.

Obviously, there is a great deal of overlap among the goals of the four groups. The government acts as the agent of society in operating the programs. As such, definitions of program success coincide in most areas for the government and for society. Similarly, individuals and employers as members of society are interested in aggregate changes as well as those directly affecting them. Likewise, the effect of programs on individuals and employers will determine in part the programs' success in terms of society. Increased employment of participants in programs is likely to improve aggregate employment, and improvement in the production of individual firms may lead to increased aggregate production.

There also may be some overlap among the goals of each of the parties. For instance, the reduction in an individual's unemployment may increase his earnings as well as decrease his feeling of dependency. Since the effects may have independent importance for the individual, we believe that all should be considered.

There may be conflicts, however, among the goals of the different parties and among various goals for a particular party. Thus, we may find that a program which improves the income of the participants is very costly to the government or that a program which is highly efficient at increasing aggregate production leads to greater inequity in the national distribution of income. These conflicts of possible program achievements raise the problem of ranking the objectives of each of the parties and of determining which party's goals have precedence. While on a theoretical level one can argue that societal objectives should be paramount, the evaluator must be a realist. He should recognize that the rewards and costs of employment-related programs to particular interested parties may play an important role in determining the size, scope, and even the existence of the program. For instance, since these programs are typically voluntary, if programs which have great potential social and government benefits do not provide sufficient benefits to participants to attract individuals into the programs, the benefits for society or government will not be realized.

Likewise, it should be noted that there are other groups, particularly pressure groups, who may be less directly affected by social programs but whose benefits also should be considered because of their political influence. Social programs as government programs must satisfy political demands as well as accomplishing economic and social good. Where the programs are located, who participates, and who manages the programs may be as important politically as what they do. Thus, the effects of employment-related programs through their positive or negative impact upon voters or campaign contributors may be extremely important to their political survival.

Even if we ignore these political considerations, employment-related programs can indirectly affect a number of people. For instance, employers will face a different labor supply and nonparticipants a different demand for their services if a program is successful in training large numbers of workers for a particular occupation. Unions might face a variety of changes if the program introduces a number of workers from minority groups or cultural backgrounds previously unrepresented in the plant. The families of participants may be adversely affected by the costs which they have to bear while the participant is in the program but may benefit if the program successfully increases family income. To conclude this discussion, one should consider the categories of persons who possibly could be affected by social programs as well as the many ways in which these programs could have an impact.

Listing of Potential Benefits From Employment-Related Programs

To facilitate the choice of impacts to be studied, we present lists of potential benefits for society, individuals, employers, and government. We feel that all employment-related programs can be judged in terms of these impacts, but simultaneously realize that each program will have a different method of reaching its objectives and will put a somewhat different emphasis on each of them. Further, because it is our hope and belief that the positive results of most employment-related programs will exceed any negative impacts which may occur, our discussion of impacts is stated in terms of benefits here and throughout this book. Finally, the list is obviously not all-inclusive. It should, however, provide many of the most important outcomes of employment-related programs. Below each benefit we present operational criteria to measure the success of an employment-related program in meeting the objective. These criteria are presented as examples of the measures which could be used. Again, the list is not meant to be all-inclusive.

A. BENEFITS FOR SOCIETY

1. Improved Equity in the Distribution of Income and Employment, Especially for Target Groups.[3]

a. *Increased Incomes.* The increase in the income of target group members relative to some stated goal such as the average income for all workers.

b. *Increased Employment.* The increase in the percentage of time in which all target group members are employed and the decrease in the percentage of time in which they are unemployed after the program, relative to the averages for all workers. A less useful measure because of seasonality and time trends would be the increase in the percentage of target group members who are employed at given times relative to a stated goal.

2. Increased National Production.

The increase in the Gross National Product (GNP) which should approximate the sum of the changes in earnings of all persons affected by the program, including persons who are not program participants.

3. Reduced Unemployment.

The decline in the average percentage of time in which persons affected by the program, including nonparticipants, are unemployed after the program. A less useful measure because of seasonality and time trend would be the change in the percentage of these persons who are unemployed at given times.

---

[3] Groups which might be considered are:
- a. Persons defined as economically disadvantaged.
- b. Members of minorities (Blacks, Native Americans, Hispanics).
- c. The handicapped (physically impaired, mentally retarded, mentally ill, alcohol and substance abusers).
- d. Groups with high unemployment rates (teenagers, the aged, ex-offenders, school dropouts).
- e. Groups receiving government benefits (welfare recipients, unemployment insurance claimants, veterans).
- f. Others (female heads of households, farmers, persons in depressed areas, the unemployed, and the underemployed).

4. Increased Social Satisfaction.

   a. *Increased Satisfaction with Social Institutions and Increased Social Participation.* The increase in participation in political activities of persons affected by the programs. The improvement in the average scores on scales of attitudes toward social institutions, such as schools, police, politics, and welfare agencies.[4]

   b. *Increased Job Satisfaction.* The improvement in average scores on job satisfaction scales.[5]

   c. *Increased Overall Satisfaction.* The increase in average scores on social indicators.[6]

5. Stable Prices.

   The stability of wages and prices in those industries and occupations in which persons affected by the program are employed relative to average changes for all wages and prices. Special attention should be given to "bottleneck" industries and occupations.

6. Reduced Antisocial Behavior.

   The reduction in the number of persons affected by the program who are arrested and convicted of crimes, who participate in riots, or who are involved in other socially unacceptable activities. Reduced recidivism rates and parole revocations for former inmates of correctional institutions might also be utilized.

---

[4] A volume of political scales should be consulted such as John P. Robinson, Jerold G. Rusk, and Kendra B. Head (26).

[5] For a compendium of such scales see John P. Robinson, Robert Athanasiou, and Kendra B. Head (24).

[6] For these measures see John P. Robinson and Phillip R. Shaver (25).

7. Reduced Dependency on Government.

   The reduction in the number of persons who receive public assistance and unemployment insurance, the amount of each received, and the proportion of time these are received. Psychological scales of dependency might also be used to examine the degree of dependency as perceived by persons whom the program affects.

8. Increased Voluntary Leisure.

   The reduction in the number of hours worked to conform with individual desires. A second measure would be the reduction in the proportion of persons affected by the program who work more than they desire. Finally, the improvement in scores on an attitude scale measuring satisfaction with leisure might be examined.

9. Improved Family Life.

   The reduction in the proportion of program-affected persons whose family lives are negatively altered (through divorce or desertion). Changes in attitudes toward other family members could also be examined.

10. Reduced Discrimination and Improved Race Relations.

    The proportion of persons affected by the program who improve their behavior toward persons of another race, ethnic group, age, and sex.

11. Improved Health.

    The average improvement in the nutritional level as measured by changes in the amount of food consumed and its protein content. The effect on health can be measured by the reduction in average number of days sick, the proportion of program-involved persons with emotional problems, and the value of health services needed by program participants.

12. Improved Housing.

    The average improvement in the quality of housing based on the Census definitions.

B. BENEFITS FOR INDIVIDUALS

1. Increased Incomes.

   The average increase in the incomes of participants. The income increase could be from either increased employment or higher levels of productivity. Separate calculations may be made for various groups of participants.

2. Reduced Unemployment.

   For various types of program participants, the reduction in the average percentage of the time after the program that they are unemployed. A less useful measure because of seasonality and time trends would be the reduction in the percentage of different types of participants who are unemployed at a given time.

3. Increased Satisfaction.

   a. *Increased Satisfaction with Work.* The average improvement in scores on job satisfaction tests by different types of program participants.

   b. *Increased Satisfaction with General Conditions.* The increase in average scores on social indicators by different types of participants.

4. Increased Social Status.

   The improvement in social and occupational status of participants with differing characteristics as measured by socioeconomic scales.

5. Increased Voluntary Leisure.

   The increase in the average number of hours when work is not sought or desired at the going wage for different groups of persons affected by the program. A second measure would be the reduction in the proportion of the groups who work more than they desire.

6. Reduced Dependency.

The reduced proportion of different participant groups who receive public assistance and unemployment insurance and the reduction in the amount of each received. The reduction in the degree of dependency as perceived by each group could also be examined. Scales of dependency might be used.

7. Improved Health.

The average improvement in the nutritional level of different types of participants as measured by changes in the amount of food consumed and the protein content in their diet. The effects on health can be measured for different groups of program participants by the reductions in their average number of days sick, the proportion with emotional problems, and the value of health services provided to them.

8. Improved Family Life.

The reduction in the proportion of program-affected persons whose family lives are negatively altered. Changes in attitudes toward other family members could also be examined.

9. Improved Housing.

The average increase in quality of housing of program participants with different characteristics based on the Census definitions.

C. Benefits for Employers

1. Jobs of Specific Employers Filled.

The proportion of participants accepting jobs in "bottleneck" industries, in occupations where workers are in short supply, and with particular employers. The number of vacancies and public Employment Service job orders filled, by industry and occupation, are a second measure.

2. Jobs in Particular Areas Filled.

The number and proportion of participants who find employment in labor shortage and/or depressed areas. The reduction in the number and the length of vacancies and unfilled job orders in Employment Service offices in these areas could also serve as a measure.

3. Improved Productivity of Particular Employers' Labor Forces.

The increase in average output per hour worked in firms which hire program participants. This might be shown by the improvement in the average level of achievement on work sampling tests for the employers' work forces. Also the change in the years of school completed for the work forces of specific employers and changes in their knowledge level as measured by achievement tests could be examined.

D. BENEFITS FOR GOVERNMENT OPERATIONS

1. Reduced Costs of Government Operations.

The reduction in the proportion of persons affected by the program who receive public assistance, unemployment insurance or other transfer payments, or who need to use the services of CETA, Vocational Rehabilitation, the Employment Service and similar agencies after the programs. This should be multiplied by the reduction in the average time spent providing services to these persons by each of the agencies involved and the cost of these services.

2. Reduced Transfer Payments.

The reduced amount of unemployment insurance, public assistance and other transfer payments received by the program participants. Changes in public assistance paid to other family members should also be measured

3. Increased Tax Revenues Through an Increased Tax Base.

The increase in the taxes paid by persons involved with the program. Separate calculations should be made for federal, state, and local taxes. The federal level should include personal income, excise, and social security taxes, and local and state tax measurements should include income, sales, and property taxes.

4. Increased Number of Persons Available for Military Service or Other Public Service.

The increase in the proportion of youth who are classified as acceptable for military service, Peace Corps, VISTA, or similar types of public service.

Discussion of Criteria

The actual measurement of the impacts is a difficult but necessary job. Consequently, some discussion of the rationale for those criteria appearing in the list above may be useful. Because economic criteria are much better defined at this time than are some of the others and therefore easier to measure, we will begin with them.

Employment-related programs are directed at improving the earnings of program participants for at least three reasons. First, as skills are improved the productive capability of society is increased (i.e., the improvement in skills of the labor force will permit society to produce more). In the various types of employment-related programs there are several techniques used in the attempt to increase productivity. The CETA Title II, Vocational Education, Vocational Rehabilitation, and WIN programs provide skill training and/or basic education in an effort to increase the productive abilities of participants. These programs also contain components to increase marketability by providing useful work habits and experience. They, along with the Employment Service, also provide labor market information, including information on where jobs are available, in order that applicants may better match their skills with the demand for them.

Next, the impact on earnings may affect the income distribution. Again, higher productivity provided by employment-related programs should lead to higher earnings—the concept of investment in human capital. Depending on who the recipients of the programs are, the earnings distribution and, subsequently, the income distribution may be altered. If, as is typical, the earnings-increasing services are provided primarily to the poor, the distribution of income may be improved.

Finally, society appears interested in improving the earnings and employment of particular groups in the society because of the belief in the "Protestant Ethic." Society appears intent on replacing welfare with work and placing a positive value on income earned as opposed to income from other sources (at least among the poor). Insofar as the employment-related programs are directed at specific groups who are poor and likely to receive transfer payments, the achievement of higher earnings and employment for them is looked upon as a benefit of the program.

How are these benefits to be measured? In the case of the change in society's production we assume it will be equal to the increment caused by the program in the productive ability of the program participants.[7] Since it is difficult to measure productive ability directly, the evaluator must rely on marginal productivity theory which says that the increment in the marginal individual's output is equal to the increment in his wage, assuming perfect competition. Thus, we can say that employment-related programs increase the output of society in an amount equal to the increment in the earnings of the program participants, given certain assumptions such as perfect knowledge and mobility and full employment.[8] This increment includes the increase in total compensation—fringe benefits as well as wage or salary payments.

---

[7] In Chapter 5, alternative possibilities will be offered.

[8] Another increment in society's output may arise from the production of program participants while they are in the program. For example, programs which provide employment in order to give work experience or on-the-job training produce goods and services as part of the process of the programs themselves. For simplicity, we label all gains which occur after the program's conclusion as benefits and all gains which occur during the program as negative costs.

It is typically very difficult, however, to get accurate measures of the value of fringe benefits on an individual basis, but an effort should be made. We usually find that fringes increase with earnings and are somewhat more substantial in union than in nonunion plants. The calculation of the increment in earnings also is before taxes and should include tax contributions by the employer on the individual's behalf. Changes in the level of transfer payments which result from the programs should *not* be included in calculations of the effects on production and capability since by definition they are payments for which no service has been performed.

Examination of the effects on income distribution and on welfare clientele will require different calculations. In discussions of aggregate production, it made no difference whose earnings were increased. However, when we consider income distribution and increasing the earning ability of the poor, we are interested in the specific increments for this group. For example, a short course in waiting room management might allow doctors to see more patients per day and increase the physicians' incomes. This increases aggregate production possibilities but affects the earnings distribution by skewing it even more to the right and does little to take individuals off of the welfare rolls. Therefore, the amount of increase in earnings for specific target groups must be measured when the distributional and welfare impacts of social programs are considered. In addition, whereas the social calculation of the production capability impact did not include transfer payments, transfers must be included in looking at the distribution of income. Likewise, concern should be with the income distribution after taxes and not before.[9]

Other types of economic impacts are the effects on government receipts and expenditures. The programs themselves require government expenditures. On the other hand, to the extent that they increase the earnings of the participants, the programs will increase the income and excise taxes that these individuals pay. Moreover, if these individuals are presently provided with social services which lead to higher earnings, future governmental expenditures on their behalf will be reduced.[10] Hence, there can be sizable impacts of employment-related programs on the government when one views the government as an economic entity which attempts to maximize the returns to its resources. This is the way that Congress and many individuals look upon government programs.

---

[9] One might also consider the effects of the program expenditures on income distribution. Some people have claimed that social programs have distributed more to the middle class than they have to the poor because it is the middle class who is paid to train the poor.

There are other economic benefits to society, the government and the individual. These, however, are indirect. For instance, to the extent that employment-related programs lead to better health, they may make for a more productive labor force. To the extent that earnings are increased, crime may be reduced thereby reducing government and societal expenditures for certain crime prevention and law enforcement agencies. There may be a reduction in the need for social services agencies if programs lead to better family life through increased earnings. Each of these indirect effects expands the production possibilities for society by either improving the resources or reducing the alternative use of resources. The measurement of these indirect economic effects, however, appears to be beyond the scope of most present studies. This is at least partially due to the fact that studies have only recently begun to look at the effects of employment-related programs on such factors as health, crime, and family life.

There are many noneconomic benefits which may result from employment-related programs, particularly for the individuals who participate in the programs. The argument is often made that these impacts are not measurable. This is not always true, as some examples will demonstrate. The possibility of improved health has already been discussed. There is a wide variety of possible measures of an individual's health. Some of these, such as days of disability, provide cardinal measures of the impact of social programs.

---

[10] In these terms, government expenditures for resources and expenditures which take the form of transfer payments are treated identically.

In the case of individual fulfillment, there is a wide variety of satisfaction and alienation scales. There are scales of social satisfaction, job satisfaction, security, and the like. In most cases, these have at least ordinal ranking and are reliable and internally consistent. For references to these scales the reader should refer to the excellent volumes by Robinson and others (24, 25, 26).

There also are political goals which may be achieved through employment-related programs. For the individuals, these may include greater exercise of political power. For Congress, it may mean reelection due to voting patterns which are favorable to the incumbent. For society it may mean maintenance of the present system by reducing the likelihood the program participants will be involved in riots. These benefits would be measured by determining the change in political behavior of the participants in the programs and the effects of the participants' actions on the political behavior of other persons.

There may also be what are considered social benefits as a result of employment-related programs. It may be thought to be socially desirable to reduce unemployment, to get people off of welfare, to have a more integrated society, to have more equal opportunity, to reduce poverty, and to improve social mobility. If these are considered as desirable benefits which may result from the programs, they need to be measured. For instance, to measure the reduction in unemployment we can look at both aggregate and individual levels. Just as earlier the increment in earnings was assumed to equal the increment in total production, we can also assume that at the margin the increase in employment of the individual will equal the increase of employment in the society. Similarly, we can aggregate reductions in welfare payments and the proportion of program participants who are moved off of welfare.

To measure the integrating effects of programs we can examine the change in housing patterns which result from the employment-related program for its participants and the change in their associations. In the area of equality of opportunity, we can examine changes in the outlook of the participants and in their opportunity for employment. In the areas of social mobility, we can look at the correlation between the fathers' and the sons' socioeconomic status. This is, of course, a long-range impact of employment-related programs, if it exists.

The Choice of Impacts and Criteria

Obviously, we believe that all potential impacts should be measured if possible. There may be time or budgetary constraints, however, which limit the number of impacts which can be examined. In these cases the evaluator should use two factors to determine which impacts to measure: 1) the expected magnitude and importance of the impacts; and 2) the ease and cost of measurement.

The assignment of priorities to specific impacts must be the ultimate responsibility of the decision maker to whom the evaluation will be delivered; that person should decide the relative importance of the measured impacts. One can argue, however, that certain impacts of employment-related programs must be included in those presented to the decision maker, such as the impacts on the employment and earnings of the participants. On the other hand, it may be argued that some impacts are likely to be so small that they are the best candidates for elimination if some impacts must be ignored. For instance, for certain programs the proportion of the participants who will have been convicted of a crime or even arrested may be so small that any impact of this program must be negligible. In this situation the omission of the antisocial behavior impact would seem justified if there were severe resource limitations and the collection of arrest data were costly. On the other hand, if the program consisted of providing employment-related services to ex-offenders, the impact could be substantial and omitting this impact would be a major error.

The ease and cost of measurement must also be considered. It should be obvious that problems will exist in accurately measuring some criteria. For instance, there will be few individuals who will admit to committing crimes or other asocial behavior except after extensive (and costly) interviewing. The data collection costs may outweigh the value of the information. Yet, as illustrated, there are many criteria which are relatively easy to measure. There are others where, with a little thought, the development of new and less expensive operational measures could be accomplished. Unfortunately, however, expedience has led potential impacts to be ignored too often, particularly for noneconomic impacts. In conclusion, it is our opinion that all potential impacts should be examined. Only where a strong and convincing case is made against an impact or criterion should it be omitted.

Exercise 2-1

Congress has given the President authorization to reorganize the executive branch of the federal government. One of the areas which seems ripe for consolidation is the provision of employment-related services to the unemployed, underemployed, and poor. Among the programs and departments which are presently providing such services are the following: CETA, the U.S. Employment Service, WIN, and the Unemployment Insurance Service in the Department of Labor; GI Bill training from the Veterans Administration; Vocational Rehabilitation, health services training, WIN, and Vocational Education in the Department of Health, Education, and Welfare; the Public Works program in the Department of Commerce; vocational training provided by the Department of Defense; programs for training and employment of offenders through the Law Enforcement Assistance Administration in the Department of Justice; and training of Indians by the Bureau of Indian Affairs in the Interior Department.

You, as an analyst in the Office of Management and Budget, are asked to list the impacts of at least four of these programs. In making up the list you should attempt to demonstrate the degree of overlap among programs in their impacts.

Exercise 2-2

Provide detailed criteria to measure the impacts listed in Exercise 2-1.

Chapter 3

EVALUATION DESIGN

When evaluating social programs, we are trying to determine the effects of these programs on the individuals who participate, the government, society, and various other parties. To do this we want to measure changes that the program has created—both economic and social changes, primarily in the lives of the participants. As discussed in the last chapter, we also use changes in the economic positions of the individuals to measure the influence of the program on society and the government. In order to determine the changes which have occurred with respect to an individual we need to know what his experience and situation have been after the treatment and what it would have been had there been no social program.

Alternative Designs

This discussion centers around various designs which may be used to measure the difference between the actual experience of an individual once he has completed an employment-related program and his expected experience in the absence of the program.[1]

---

[1] In a primer such as this there are obvious limitations on our ability to cover the topic of experimental designs. We strongly urge the reader to examine some of the theoretical literature dealing with experimental design. We particularly recommend Campbell and Stanley (12).

The Case Study Design. Probably the simplest design is the case study. Here, a treatment is introduced and the participants are observed after receiving the treatment. Then, based on a guess as to what would have happened to the participants had there been no program, a judgment is made as to whether or not the treatment improved the lot of the participants. This kind of design is quite common in the evaluation of social programs. Many of the manpower programs begun in the 1960s grew out of experimental programs which had no evaluation other than a case study. We have subsequently found that in many instances the expanded program has been relatively unsuccessful.

There is a variety of major problems with the internal validity (the accuracy of the findings for the group studied) of the case study design which caused this unfortunate result. First, there are events outside the control of the evaluator which influence the observations that he makes.[2] To give a concrete example, employment-related program placements depend for their success largely on the labor market for which they provide participants. During the middle 1960s the labor market improved considerably, and it appeared that any program which provided warm bodies could find at least some job for these bodies. The improving labor market would lead to individuals getting better jobs than one might otherwise suppose based on the experience of the early 60s. When the economy slowed down in the early 1970s, the opposite situation occurred. Thus, there is a very great opportunity for the *post hoc ergo propter hoc* fallacy to exist. The evaluator implicitly says "In my opinion, the participant was better after the treatment; therefore, the treatment caused the participant to get better."

Second, the mere passage of time will influence the expected labor market experience and the attitudes of most people.[3] This is particularly true for the very young. For instance, teenagers have extremely high unemployment rates which tend to go down as they become older and more acceptable to employers. Consequently, a program that takes in a youth when he is 18 and turns him out when he is 21 will find that his employment possibilities have improved after completing the program. Some programs, such as the Neighborhood Youth Corps, have been characterized as merely "aging vats" for the young.

---

[2] Campbell and Stanley (12) call this history.

[3] Campbell and Stanley (12) call this maturation.

Another problem with the case study is that it is difficult to tell if there is a selection problem, particularly in the case of small-scale demonstration projects. There is the distinct possibility that a small-scale project will be able to "cream"—take only the better prospects for employment success. It is not surprising, then, that after having received the treatment of the program, these small-scale project participants are found to do better than the average person.

Before-and-After Designs. While the case study approach has been used to make many actual policy decisions, employment-related program evaluation has tended to follow a slightly more sophisticated design. The before-and-after design has been used in a number of government publications describing the benefits of social programs. The selection problem is removed since a comparison is being made with the same individuals, i.e., if they were above average in employability after the program, they were presumably above average before it. The other problems still remain, however. In fact, they become more serious for we are now considering a longer period of time in which events and maturation take place.

Probably the most serious problem in using observations before and after program participation for the evaluation of social programs is the problem of "regression toward the mean." It is an observed phenomenon that in large populations which are examined over time, those values at the extremes tend to move towards the middle. To give a specific example, if we look at a cross section of hours worked by the labor force, we find at one extreme individuals who are employed for a fantastic 5,200 hours a year. At the other extreme are individuals who are unemployed for the entire year. If we make a cross-sectional analysis of hours employed in the following year we find the same situation. Some individuals are working more than full time all the time, some individuals are totally unemployed. If we follow the individuals in either of these categories, from one year to the next, however, we will find the individuals who are working the most hours in the first year have a greater tendency to reduce their hours of work than does the population as a whole, so that their average number of hours worked will fall relative to the population. This is not unreasonable, for they have more hours to reduce. At the other extreme, the individuals who are totally unemployed in the first year will tend to have greater than average increases in employment during the second year. Again, the only way they can move is up.

The problem in terms of the evaluation of employment-related programs, then, is this: the individual who is eligible to enter these programs is typically at the bottom extreme of the labor force spectrum. Most of the programs require that the individual be poor, unemployed, underemployed, on welfare, handicapped, or otherwise in need in order to qualify for program assistance. These types of individuals are the ones who would be expected subsequently to have higher than average increases in their earnings and employment because of regression toward the mean. Thus, when they do show increases in these variables, it is difficult, if not impossible, to say what amount of the increases is due to the program and what amount is due to regression toward the mean.

We consequently have three major problems with simple before-and-after evaluations: the influence of extraneous events, the mere passage of time, and regression toward the mean.[4] An alternative formulation of before-and-after studies attempts to get around some of these problems. This method, labeled the interrupted time series, makes repeated observations before the program and then repeated observations after the program. From the repeated observations, some indication of the general trend caused by maturation should become evident. This method should also show some of the regression toward the mean by evening out the cycle through which the individual is going if a number of observations over a considerable period of time can be made. However, such repeated observations are not possible or useful in the cases of new entrants to the labor force who have no past experience to judge and reentrants to the labor force who have been out of the labor force for long periods of time. Moreover, this procedure may require continuing observations of the individuals over a long period before they are allowed to go into the program, which is obviously an expensive as well as time-consuming proposition and as such probably not practical in many instances.

---

[4] There are other less important problems as well. See Campbell and Stanley (12).

Another alternative is to try to predict the expected before-to-after change through the use of multiple regression analysis. This procedure uses such independent variables as age to cover maturation, and growth in the economy to try to account for the problem of intervening events. The assumption is then made that the predicted earnings of the individual resulting from the program are net of these influences. This method, however, will not handle the regression toward the mean problem. Furthermore, it is very difficult to arrive at a regression model which accurately specifies the relationships of such variables as earnings and employment with explanatory variables. Studies using cross-sectional data which explain 20-30 percent of the variation in earnings or employment are considered to be quite good. This leaves at least 70 percent of the variation unexplained.

Comparison Group Designs.[5] To get around a number of the problems involved in the before-and-after comparisons, evaluations of social programs increasingly have attempted to use a comparison group to represent the expected experience of the participants in the program in the absence of their participation. The key to the use of comparison groups lies in how well they represent the experience that the program participants would have had in the absence of the program. If they do not closely approximate the expected behavior and experience, all of the problems previously discussed come into play. The easiest way to see this is to view all program participants as black youth and the comparison group as white prime age males, and then look at the expected labor market outcomes. For instance, we know that the percentage reduction in unemployment among blacks is likely to be considerably greater in an economic upswing and their loss of employment considerably greater in an economic downswing. Depending upon which stage of the business cycle we examine, we would get different results from the program solely as the result of the choice of the comparison group. Similarly, we know that black youth would gain considerably more in employment from the passage of time. Therefore, particularly with a before-and-after comparison, the program group (the black youth) would show greater gains than would the comparison group of white prime age males, regardless of what the program did for the individuals involved. Finally, in terms of regression toward the mean, the youths who participate in the program are probably among the most disadvantaged in society. They have nowhere to go but up, whereas the white prime age males probably are at their peaks and can expect declines as they grow older. The differences between the treatment group and the comparison groups are vital.[6]

---

[5] We label as "comparison group" any group whose expected characteristics and labor market outcomes might not be identical to those of the participants in the absence of program participation. A group which was randomly selected from the same population as the participants but not allowed to participate we label a control group since, if the numbers are large and the selection is truly random, they should be identical to the participants.

Several sources have been used to secure members of comparison groups. First, evaluations of on-going programs have used individuals who had been applicants to the referral agency (e.g., the Employment Service) at the same time as the program occurred but who did not go into the program. This is not necessarily a good comparison group because these individuals could differ from the program participants in the following respects: 1) they did not go into the program because they had offers of employment or possibilities of employment which they considered to be better than they would have after completing the program; 2) they did not go into the program because they did not feel that they had the qualifications necessary to complete it; 3) they were excluded from the program because they did not meet the entrance qualifications; and 4) they lacked the motivation to enter the program. One could argue in the first case that the comparison group is superior to the program participants and in the other three cases that they are inferior (in terms of expected labor market outcomes). A post-program observation of both groups should definitely expect to find some differences in their experiences, attitudes, and behavior because of these four reasons. If a before-and-after design were used, it would be very difficult to say what interactions of differences between the groups and history, maturation and regression toward the mean would be expected—but they certainly could exist because of the selection process.

---

[6] To use Campbell's and Stanley's (12) terminology, there are interactions between selection and history, maturation, and regression toward the mean.

A similar type of situation exists for another common comparison group—one made up of individuals who applied to the program and who were deemed qualified but who did not enter it. This comparison group presumably is more able than that described above which includes persons who did not meet the entrance qualifications. But it still may not be comparable since the problems of self-selection (items 1, 2, and 4 above) or program selection, if the program took the most able of the qualified, remain.

Other studies have selected as comparison groups people in the same neighborhood who were in the same condition as the participants before the program. This group is one step removed from the applicants in that they presumably did not apply for the program, perhaps because they lacked the motivation to do so. Another similar group used particularly in studies of vocational education and youth programs are individuals whose names are taken from the files of the high schools attended by the participants but who do not enter the particular program. Again, there are problems of motivation. The questions arise as to what factors not on the file card differentiate these individuals from those who entered the program and whether or not these factors would in turn lead to differences in the outcomes which are to be observed.

In all of these cases of comparison groups, one can attempt to match them with the participant group through statistical control. To the extent that the two groups differ in identifiable characteristics which are thought to affect the outcome measures and which are quantifiable, the selection procedure or regression techniques can be used to take some account of these differences. A problem arises, in that our models of what causes the outcomes are not well specified and complete and are measured with error so that the regression analyses will not measure all of the differences.[7] Consequently, the matching process does not guarantee that the participants and comparison group members come from the same population. One person may be at the top of a low distribution and the other at the bottom of an overlapping, higher distribution. If this is the case, they will regress to different means.

Also, there is a more basic problem in that we are unable to measure many of the variables which we believe affect the outcomes of social programs. For instance, motivation must be an important factor, but it is difficult to measure on a before-and-after basis—and impossible on a retrospective basis. For these reasons, although they are desirable, statistical control methods are unlikely to solve the problems of the comparison group differences mentioned above.[8]

Random Control Groups. The answer to these problems is the random assignment of eligible persons to participation or to a control group. All of the persons who are qualified to enter the program must be contacted immediately before the start of the program to find out if they are still interested in entering it. The group still wishing to be considered would then be split randomly with only one group assigned to actually enter the program. The second group would be given the regular services, if any, normally available to them. It is possible that this method could yield too small a number in the study who possess a particular characteristic which is important for analysis. If so, the group wishing to enter the program should be stratified by this characteristic. Then differing sampling proportions would be used from each of the strata. The sampling from each strata, however, would be random.

---

[7] A discussion of these factors may be found in Director (13). Director also shows in a review of the studies of manpower programs that the studies' findings appear dependent on whether or not the comparison groups are superior or inferior to the participants.

[8] For other approaches see Cain (8) and Heckman (18).

Random assignment procedures have the outstanding advantage of being statistically sound when they are applied to large groups. Known probabilities can be given to chance differences in the success of the two groups.[9] This would not be true of other means for selecting comparison groups. Thus, any differences between the participants and control group observed in post-program situations can be attributed to the program with known confidence levels.

Practical Problems of Random Assignment. Several problems are involved in the random selection of a control group. The first is the reluctance of the operating agencies to exclude fully qualified persons from receiving the services of the agencies. Such an exclusion is felt by the program operators to be analogous to a doctor not treating a patient who has a disease when a drug is available. To carry the analogy further, however, we demand in medicine that drugs be tested before they are administered so that their effects are known. This testing is done through just the process recommended here. Persons with the disease are divided randomly. One group receives the drug and the other group receives placebos. This is the only scientific manner to test the actual effects of the drug. In medicine, it is recognized that it could be much more disastrous to administer a drug which has no effects or which has dangerous side effects because of an untested belief that the drug is beneficial than to test the drug properly and, if it is ineffective, substitute a better alternative. The same can be said about social programs.

Implicit in random assignment, however, are decisions about who will go into the programs which may be different from those which would have been made in the absence of the evaluation. While this might not greatly concern the evaluator, it doubtlessly will be of great importance to the operating staff of the programs which are being evaluated. The staff may view the random selection procedure as basically denying the criteria used for program admission so that the evaluation is reducing the efficiency of the program and, by so doing, harming the participants in the program. To them, random assignment is effectively denying services to persons who have sought services in the belief that these services would improve their position. Furthermore, it is only natural that the program staff who must account for the program and their own work in that program should believe that it does help the participants. (It would be a sorry situation if the operating staff of social programs did not believe that they and their programs were accomplishing good.) For one critic of random assignment the question boils down to: "When one realizes that . . . programs have been established primarily to assist usually discouraged and disadvantaged individuals to better their lot rather than for research purposes, how can one deny them service that they need and applied for? Moreover, how can one continue to deny them service over a period of time despite their urgent need and interest?"[10]

---

[9] For instance we can say that 95 times out of 100 a difference as large as we observe is not due to chance.

There are several possible answers to this question. First, and probably the most important, is the answering question "How can you be sure that the social program in question actually does assist the participants to better their lot?" Social programs have come into being and grown very quickly in the last fifteen years; yet, they have been subjected to constant revision. These changes presumably occurred because the programs were either not operating to the benefit of the participants or needed considerable improvement. We have seen major changes in the structure of social programs. For instance, CETA replaced a variety of categorical manpower programs and shifted decision making power from federal to local officials. Programs have changed in size—the Job Corps was decreased in size substantially during 1969-75 and was increased after 1977. Programs have changed clientele—the psychogenically disabled were first added to and then dropped from Vocational Rehabilitation eligibility. Programs have changed content and emphasis—vocational education amendments have repeatedly downplayed agricultural and home economics training in recent years. Presumably all of these changes occurred because it was believed that individuals served by these programs could be better served by other programs or no programs at all. These changes did not come about, however, until literally hundreds of thousands of individuals had gone through the programs because typically there was a delay of several years between the implementation of a program and the gathering of data to show that the program was not operating properly. These individuals did not receive services which were subsequently determined to have been more appropriate for them and they even could have been harmed by the services that they did receive.[11] What we have here is a situation somewhat similar to the doctor who presents a quack "miracle cure" for cancer to his patients: 1) some benefit from it; 2) others survive it; 3) some, because they have taken the "cure," do not seek any other means of treatment and subsequently are worse off; and 4) others die from the treatment itself. One need only postulate the third alternative to see the risks that are inherent in social programs not being properly evaluated.

---

[10] Abraham Stahler in Borus (6), p. 68.

On strictly practical grounds there are seldom enough positions in a social program to satisfy the universe of need. Almost always there is a rationing of program openings. Ideally, the positions are assigned to those most needing the program services. Usually, however, assignment is on the basis of some other, less beneficial, criterion such as on a first come-first served basis for those who qualify for the program or to those most likely to succeed in the program ("creaming"). The random assignment technique will not reduce the amount of services provided nor the eligibility of the participants. It will merely change which qualified people receive the services and may more equitably distribute the benefits of the program by giving qualified individuals an equal chance to participate.

---

[11] One could argue that many of the individuals who went through the early programs were harmed by them. The individuals who took training and subsequently did not get jobs would be a good example.

Another problem with the random assignment technique is that it requires more extensive selection than would normally occur. Under ordinary conditions it is only necessary to interview, test, and refer persons to the program until all slots are filled. With random assignment, enough additional persons must be found who also qualify for the program so that a control group of adequate size can be selected. Since the projects selected for evaluation will form only a very small part of each program, the increase in cost will be relatively small. It should be viewed as a cost of the research which will be well repaid in terms of the accuracy of the results.

If the control group and the experimental group are identified at the start of the project, efforts can be made to keep track of these persons. This will improve the response rate for the evaluation and reduce the costs of follow-up. The choice of retrospective comparison groups usually involves trying to contact persons with whom there has been no link for many months. Many cannot be located, and a great deal of time and effort is spent trying to locate the others. Thus, the use of the random assignment technique may lead to a lower total cost for the evaluation. Finally, while the random assignment technique requires a longer time period from the start to the end of the evaluation than would a retrospective study, the two techniques will present their findings at the same time for a new program or a changing program.[12]

Theoretical Problems with Random Assignment. Unfortunately, while the randomized assignment designs are accurate assessments of the program for the individuals who have participated in it and who have been studied, like all other designs there can be a variety of external validity problems (i.e., there are problems in generalizing beyond the particular group which has been studied). These same problems apply to the nonrandom designs discussed earlier. They include the following. 1) The pretest observation and the treatment itself may interact. For instance, if before the program I am tested on my reading ability, I may well concentrate on reading when I get into the course. My instructors may do likewise. Thus the fact that the pretest observation has been made may influence my subsequent behavior in the experiment. 2) Another similar problem, which Campbell and Stanley (12) call "reactive arrangements" and which is more generally known as the Hawthorne effect, is probably more important and more difficult to handle. This occurs when the individuals realize that they are receiving special treatment and so perform differently simply because they are participating in an experiment. Similarly, and possibly more important for social programs, the providers of services (the staff), knowing that they are being evaluated, may perform differently. 3) The selection of the groups to be studied may be such that they are not truly typical of the general population. For instance, a small pilot program, simply because it is small, may be more selective in choosing participants than a national program might be. A more serious problem is a frequent tendency to generalize findings to other groups beyond those which have been studied. For instance, the WIN program currently has a major voluntary component, and, through an experimental design, accurate measurement of the program's effects for these individuals can be made. These measurements, however, may not apply to the same program when administered on a compulsory basis under a work test. 4) Accurate assessments may be made given a certain history of events, but these may not hold for other conditions. For example, it is possible to measure accurately the effects of a program that provides on-the-job training in a period of prosperity which will not be an accurate measurement for a period of decline since the program is highly dependent on business conditions. 5) Finally, in a somewhat similar vein, there is the criticism that evaluations often examine as a single entity a social program which consists of a number of different services. These services need not be consistent from one location to the next. Conceivably, we can have a situation where a program worked in one city because of a charismatic director, but it may not be possible to recreate his charisma in leaders in a number of other cities. Consequently, the program may not be as successful there. Yet, the evaluations do not examine the components of the program and the reasons for its success and failure. The program is treated as a "black box" that is not opened to examine the contents.

---

[12] The advantages of random assignment have been recognized in several recent experimental projects conducted with funds from the Office of Research and Development, Employment and Training Administration, U.S. Department of Labor. The most notable of these is the Transitional Aid Research Project and the Supported Employment Program. In both programs, hundreds of people have been randomly assigned to experimental programs without incident.

The reader interested in further arguments for and against random assignment should see Boruch (3), Campbell and Boruch (10), and Campbell and Erlebacher (11).

These problems can be handled in a variety of ways. The pretest interacting with the treatment is probably the easiest to handle. One need only exclude the pretest. The use of the "post test only design" may also be useful in removing the Hawthorne effect for the participants, since they will not know they are being evaluated. In terms of the program staff, one can try not to notify them that an evaluation is taking place. Of course, this is usually difficult to do. Alternatively, one can evaluate a sufficiently large group of projects over a long enough period so that only a small proportion of any given program is to be sampled. This makes it extremely difficult for the program staff to decide who gets special treatment. For the problem of self-selection and creaming, one can attempt to diversify the types of program participants to be studied.[13] One can repeatedly evaluate so that the program is studied under many different conditions to account for the problem of generalizing beyond one set of economic conditions. Finally, if there is a "black box" problem, one can attempt to observe the operation of the program or to disassemble the program into its components and then evaluate the effectiveness of each of these.

Another possible problem with random assignment comes about because nonselection for a social program can conceivably have a negative effect on outcomes for the control group. If services are refused and the individual is told that he is unqualified for the program, this may be one more of a set of disappointments to him which reduces his desire, motivation, and ability to function. If this is the case, then the selection process itself must be considered as externally invalidating the results by giving a positive bias to the measured success of the program. Therefore, one must devise some kind of assignment process that is neutral in its effect on those who are to serve as the control group. Unfortunately, it is very difficult to design a placebo social program. Since social programs are widely touted in the press and in various outreach functions, it is difficult to tell any individual that he did not really want to enter the program because we are not sure it would be good for him. Whereas some social programs may not do very much for an individual, it is hard to design one which will have no positive or negative effects and yet not be viewed as a sham by the participants. This problem is a major issue which is as yet unresolved in program evaluation. The answer probably lies in honesty. The potential clients should be told that some will not be enrolled and that the decision will be made randomly.

---

[13] If the criteria for self-selection and creaming are known, then these can be included either directly or with an instrumental variable in the model. Such criteria are seldom known, however.

Finally, we come to the problem of a lack of independence between the treatment group and the control group. A program may be designed to increase employment of participants in a small labor market area where there is only a limited number of job openings. In the absence of the program, the jobs would be distributed randomly among the unemployed, whereas when the program does exist all job openings are filled by the program participants and none go to the control group. In this situation the control group is not a good proxy for the experience of the program participants in the absence of the program. As postulated, in the absence of the program there would be random distribution of the jobs, and some of the people who subsequently participated in the program would have been hired. With the program, however, no one in the control group will be hired. In this case there would be an overstatement of the incremental benefits of the program.

The lack of independence may also work in the opposite direction if the benefits of program participation are transferred from the participants to the control group. Such a situation is most likely to occur when there is close contact between the participants and control group members. This may occur when they all come from a small area and when the program provides information which is easy to convey, such as how to write a resume or the name of an employer who is hiring. In these circumstances there is an understatement of program benefits because some of the benefits of the program will accrue to the control group. The purpose of the control group is to show what would have happened in the absence of any program; if the control group is influenced in any way by the existence of the program, it does not truly reflect the experience of the participants if there were no program.

These problems are not easy to resolve. One alternative is to conduct the evaluation while the program is still small, relative to the labor market area. (If the evaluated program is turning out a hundred people in New York City, its impact on the job market will be extremely small.) The evaluator, however, should be aware of the threat to external validity due to the lack of independence of the experimental and control groups, and should attempt to prevent the participants from becoming such a large fraction of the total to be hired that they overshadow the control group.

Alternatives to Random Assignment. In the situation where random assignment is not allowed because it is believed to interfere with the enforcement of strict eligibility requirements, Campbell and Stanley (12) offer an alternative—the regression-discontinuity design. Program operators are asked to specify all selection criteria that they wish to use and then to rank all applicants by these criteria. They then follow these rankings in selecting individuals for admission.[14] For instance, all individuals with incomes below $3,000 may be admitted and all those above excluded or all persons scoring above a grade of 50 on an aptitude test may be entered and those at or below 50 excluded. The postprogram outcomes are then regressed separately on the rankings of the two groups (e.g., incomes or test scores). If significant differences between the two groups appear at the cut-off point ($3,000 or a grade of 50), this would indicate that the program had an impact, at least at that point.

---

[14] Exceptions who are admitted although they do not meet the criteria and persons excluded even though they meet the criteria are identified and not considered in the evaluation.

There are several obvious virtues to this procedure. It allows the program operators freedom to make all program assignments, it does not involve any unusual effort on their part, and it can be adapted to many types of programs. However, it also has some shortcomings. First, there should be no natural discontinuities in the outcome measures. Second, the effect of the program is only measured at one point—the technique does not allow a determination of whether the program would be equally effective for persons with incomes of $10,000 or test scores of 25. Third, if the relationship between the selection variable and the dependent variable is not properly defined (e.g., a linear regression is used when the data are not linear), incorrect estimates may occur. Finally the technique does not solve any of the external validity problems discussed earlier.[15]

Another alternative is often proposed if the evaluation seeks only to determine which of several programs is preferable. Persons qualified and interested in participating would be randomly assigned to one of the programs. Recruitment would not exceed the number of program slots and a control group that does not enter any of the programs would not be necessary. Persons in each program would serve as control group members for comparison with the participants in other programs.

Notwithstanding the advantages of this type of analysis, there is a problem in that it gives the increments in the benefits of one program over another as opposed to the increment in the cost. This does not yield what it is really necessary to know: some measure of the ratios of the total benefits to the total costs for the two programs. For instance, we may find that Program A may yield a present value of $500 more in lifetime earnings for an additional cost of $300 when compared to Program B, which argues for investment in Program A rather than Program B. If, however, Program B is providing a present value of $2,000 in lifetime earnings at a cost of $100 (as compared to no program) and Program A is providing $2,500 worth of lifetime return for a $400 investment, we can get a return that is on the average more than three times greater from Program B and obviously should invest in it. Only in the situation where we know the cost and benefits of one program as opposed to no program or in that situation where the two programs have highly comparable costs can accurate interprogram comparisons be made.

---

[15] The reader is strongly urged to read Campbell (9) and Boruch and Riecken (4), pp. 87-116 for more discussion of the regression-discontinuity design. For instance, it might be argued that if the relationships on both sides of the discontinuity are properly described and identical, then one can estimate the impact of the program at points other than the discontinuity.

The Choice of an Experimental Group

Of the persons assigned to the program, some will drop out before they ever enter the program, others will enter the program but will leave before it is completed, and finally, there will be a group of completers. Among the control or comparison group, if there is one, no one will participate in the program. Some studies have argued that only those individuals who completed the program or only those people who completed the program and made use of it should be included in the calculation of the program's benefits. Such a procedure must make two assumptions to be correct: 1) the individuals who did not go into the program and the individuals who dropped out of the program were totally unaffected by it, and 2) these individuals did not differ in their expected postprogram experience from the participants who completed the program. Such assumptions probably are not warranted.

It is quite possible that those individuals who dropped out of the program gained some knowledge or work experience while they participated which might subsequently be of use in the labor market. On the other hand, while in the program they may have lost time searching for a job and may have missed job opportunities or they may have lost seniority if they had foregone a job in order to enter the program. There may also be a stigma attached to them in that employers might consider the dropouts unstable because they did not complete the program and not want to hire them. The individuals who did not enter the program or who dropped out of it also may feel rejection or lack of ability which could subsequently affect their functioning in the labor market as well as in their lives. Similarly, the program may affect all the people who complete it, not only those who make use of it. Beneficial or negative changes in attitudes may occur among program participants even if they do not appear to be affected in their employment or some similar outcome. If any of these effects on nonentrants, dropouts, or nonusers occurred, the effects of the program on *all* participants must be examined if a full accounting is to be achieved.[16]

With regard to the second assumption, the control group is used to represent the aggregate experience of all persons who were selected for the program. Unless the participants are a homogeneous group, which earlier we argued was unlikely, it will be necessary to separate the control group into segments corresponding to completers, dropouts, and "no-shows" in order to make comparisons. Such a division will be very difficult since identification of the factors which led individual participants to complete, drop out, or not enter will be required. As discussed previously, our ability to measure and model such factors is very limited. For instance, some studies of employment-related programs show higher postprogram earnings for persons who remained in the program longer. There is no good test to determine whether this was due to their learning more from longer participation or whether greater motivation caused them both to remain longer and to have higher earnings.

---

[16] This does not mean that no distinctions should be made between different types of program participants. The analysis should seek to determine the differences which exist between completers and dropouts in order to determine the necessity of reducing the proportion of dropouts or the possibility of shortening courses.

The Timing of Impact Measurement

Since the purpose of evaluation is ultimately to affect policy relating to the program's operation, there is a variety of pressures which move toward early measurement of the impacts and costs of the program. Policy makers and politicians, anxious for pilot programs which appear useful to be expanded to nationwide status, often do not want to wait for the results of the evaluation before proceeding. Program managers want to know which of the alternative funding possibilities are most profitable so that they can make their annual allocations. Program operators want results which will justify their program to the program managers and will permit them to alter their program to make it operate more efficiently.

Countering these pressures for immediate impact measurement are the following concerns. Pilot programs need time to work out the "kinks" in their operations. The program operators must: 1) establish a series of procedures for treating clients; 2) hire and train staff, weeding out those staff members who cannot perform; 3) announce the program and attract suitable clients; and 4) operationalize the program. Undoubtedly, these steps will require time before the program can operate efficiently. If the program is evaluated before this has happened, there will probably be a downward bias in the estimate of the program's impact.[17] In addition, it is necessary for sufficient time to pass after the program treatment for the transitory effects of the program to be dissipated. Finally, seasonal factors should be removed from the data by allowing at least a full year to pass after treatment before impact measurement.[18]

For these reasons it makes sense to delay the evaluation of programs until they have operated for at least six months. Then measurements of program impact should not occur until one year after the participants have left the program. Evaluations should also be made at three or five year intervals after the program participants have terminated to find the longer run impacts. In all cases, care should be taken to insure that the data are collected for identical time periods for both the experimental and control groups. Otherwise, problems with cyclical fluctuations may arise. Variables such as earnings and employment should be measured for the entire postprogram period as well as for the individual years. Such measurements will demonstrate the total effect of the program and changes in program effects over time.

---

[17] There can be an upward bias, however, if the program "creams" its clients at the beginning.

[18] This is seen by using a farming-related training course as an example. If the measurement is made on weeks employed during a three-month period ending in September, considerably different results will be obtained than if the measurement is made for three months ending in February.

The Choice of Retrospective or Concurrent Data Collection

One can identify participants and comparison group members associated with an earlier program and gather information retrospectively. Alternatively, data collection can occur by identifying a group of future program participants and control or comparison group members and then gathering information from or for them over succeeding periods. There are merits to both procedures.

Retrospective studies provide results faster; they do not require waiting for the postprogram period to occur before measurements are made. This shorter elapsed time from the decision to conduct the evaluation to the presentation of its findings is the reason most evaluations have been conducted in this manner, although another argument in its favor is that there will be no "Hawthorne effect" since the individuals and staff do not know that data will subsequently be gathered about them. Further, retrospective data gathering does not require that the evaluation be built directly into the program's operation, which makes advance planning unnecessary.

Offsetting these factors, concurrent data collection offers several important benefits. First, it is impossible to have control groups in a retrospective study because random assignment is impossible after the program has ended. It should also be noted that it will be difficult to implement the regression discontinuity design on a retrospective basis as the selection criteria will usually not be made explicit in normal program operations. Therefore, retrospective studies must construct ad hoc comparison groups from whatever records exist. This usually involves substantial difficulty since programs seldom maintain information on nonparticipants and to the extent that these lists are incomplete or inaccurate, biases may be introduced. Next, data gathered on a retrospective basis are much more likely to involve response errors. The longer the period to be covered, the more likely will be memory lapses. Also, studies show that accurate attitudinal measures are very difficult to collect on a retrospective basis because perceptions of past attitudes are altered by intervening events. Finally, as argued earlier, retrospective studies are more likely to have inadequate response rates. It is more difficult and costly to try to locate individuals after contact has been severed than to maintain contact with individuals.

We find the set of arguments for concurrent studies to be much the stronger. In our opinion, the opportunity to use a control group or the regression discontinuity design outweighs the time consideration. We believe that valid data are much better than early, but inaccurate, information. Also, the greater ease and lower cost of concurrent data collection have much appeal. Therefore, we strongly urge its use.

Collecting Data on the Impacts

Direct contact has been the basic method used to collect information about the program participants and control group members. Special studies have usually used personal interviews, while the government follow-up system has relied primarily on mail questionnaires and telephone interviews. The problems with direct contact are well known: besides being time consuming and costly, response rates on personal interviews seldom are above 80 percent and those on mail and telephone surveys are considerably lower.[19] Yet, for many variables which are affected by social programs, direct contact may be the only source of information. For instance, noneconomic data on program participants is not normally collected by existing data collection processes.

---

[19] For an excellent review of these issues we suggest the reader see Weiss in Struening and Guttentag (35), pp. 355-95.

The random selection technique presented earlier should reduce some of the problems associated with contacting individuals. Since the individuals for whom data will subsequently be collected are known from the start of the program, special efforts can be made to remain in contact with them during and after the program. For instance, sample members may be given stamped cards to report all changes of address, the names of relatives and others who would know the location of the respondent could be secured, and the need for follow-up information could be impressed on the sample members. This should aid in achieving a higher response rate and reducing the costs of the survey, particularly when contact needs to be maintained for a period of at least one year after the program.

Because of the variety of information on dependent variables needed at the end of the first year following the program, personal contact appears to be the best technique to use for the first evaluation of programs. Subsequent evaluations may consider the use of existing records.[20] Some of the data sources which may be used include the following.

Social Security Administration Data. For earnings data, direct contact may not be necessary. The Social Security Administration has earnings information by quarter for all covered employment. Several problems are involved in the use of these data, however. First, coverage under the Social Security Act is not universal. Of particular importance to the persons likely to participate in social programs is the exclusion of federal employees, employees of government and nonprofit organizations whose employers did not volunteer for coverage, and railroad employees. In addition, wages paid to domestics and on "odd jobs" often may not have been reported either because the worker did not earn the $50 per quarter minimum from the employer or because both the employer and employee were ignorant of the law. Coverage excludes approximately 10 percent of all employment.

---

[20] A problem becoming increasingly more important is the reluctance of various government agencies subject to federal and state "privacy acts" to provide data for identifiable individuals. Without the ability to link records for individuals, many of the economies associated with the use of existing data sets are lost.

Next, not all earnings are taxable; employers are required to report only the taxable amounts.[21] Consequently, individuals with reported earnings equal to the taxable limits in fact usually had greater earnings, requiring the reported earnings to be extrapolated. Some individuals, though, have reported earnings in excess of the limits. In a few cases these higher reported earnings could have been caused by "overreporting" by employers (some employers choose to report earnings beyond those required). In this case the relationship of the recorded to actual earnings would depend on how much the employer chose to report. More often, recorded earnings above the maximum are due to employment in more than one job during the year covered by social security. In this situation, the individual may not have earned the taxable limit on either of the jobs or he may have earned the taxable limit on one or more of the jobs. To determine which was the case requires visual checks of individual records which increases the cost and time involved. The general policy of the Social Security Administration is to make data resources available within recognizable limits. Such limits include safeguarding the confidentiality of information on individuals so that data will not be provided on specific individuals.[22]

Finally, there is a time lag between the end of the calendar year and when the data are posted. Approximately 98 percent of nonfarm and 90 percent of farm wages are posted by the end of the following October. However, only 65 percent of the self-employment records are posted then and it takes another six months before 90 percent are posted. Thus, complete data are not available until nearly one and a half years after the end of the year being examined. In addition, more time may be required by the Social Security Administration before they can handle the job. For long term follow-ups of individuals, however, this may be the only source of information available. It may not be possible to locate participants many years after a program—the mobility of individuals in social programs is high. The Social Security Administration, on the other hand, always has complete and relatively accurate data.[23]

---

[21] For the years 1959 to 1966, only the first $4,800 of covered earnings were taxable. In subsequent years the taxable limit was raised; $6,600 in 1968, $9,000 in 1972, $10,800 in 1973, $13,200 in 1974, $14,100 in 1975, $15,300 in 1976, $16,500 in 1977, $17,700 in 1978, $22,900 in 1979, and $25,900 in 1980.

[22] Data will be provided in tabular form including the sum of squares and cross products matrices for the entire sample or subgroups of the sample (containing more than four persons) as designated by the researcher.

Tax Data. The Internal Revenue Service (IRS) may provide earnings data as contained on the 1040 series forms. If coded variables of the individual's background characteristics and his social security number are sent to the IRS, the Service may provide tabular summaries on individual income data, provided that three or more individuals share the same coded characteristics and that the identifying social security numbers are omitted. Data are usually available for a given year by the late summer of the following year.

The main obstacle to effective utilization of this procedure is the lack of identification of the source of earnings, an important consideration for married persons who file a joint return. Combined earnings of husband and wife cannot be separated. Thus, these data are of greatest use for evaluating youth programs where most participants file single returns. Arrangements for this service are made through the Assistant Commissioner for Research and Planning of the IRS. Since access to the main files is limited, priorities are established by the IRS.

Unemployment Insurance Wage Data. Thirty-seven states and the District of Columbia require all employers covered by their unemployment insurance laws to report the quarterly earnings of all covered employees. All earnings are reported; there is no limit to the amount. Also, posting of these data is usually complete within three months of the end of a quarter. Consequently, unemployment insurance data do not suffer from two of the major faults of social security data.

---

[23] Additional information on the use of Social Security Administration data may be obtained from their Office of Research and Statistics in Baltimore.

There are other problems, though. The states which do not collect these data include many of the major industrial states. Even in states which do collect the data, coverage is not extended to the self-employed, persons who work for immediate relatives, employees of small firms, and some employees of nonprofit units. Separate systems exist for the armed forces, federal government, and railroad personnel. Furthermore, because the system is state-based, there is no way to distinguish individuals who have left the state from those who are no longer in covered employment or who have left the labor force. In all three cases there would be no record of earnings. This is particularly important in labor markets which cross state boundaries, such as New York City and Cincinnati. Finally, the states usually hold earnings records for no more than about two years.

Thus the unemployment insurance data will be most useful for short term follow-ups in industrial areas. In these cases it will provide accurate data quickly for most of the sample. For longer periods when the person is more likely to have left the state, social security data appear to be a better alternative because they are national.

Sample Size

The problem of how many participants and control or comparison group members should be examined is always a very difficult one. The usual rule of thumb is to include as many as the budget will allow. At times more sophisticated justifications must be provided. In these situations the best strategy is to contact a sampling expert. If one is not available, the simplest calculation is dependent on the dispersion of the variable being studied and the level of precision desired. The formula for the standard error of the difference between means for the participants and control or comparison group members is:

$$S_{\bar{X}} = \sigma / \sqrt{N(\pi)(1-\pi)}$$

where: σ is the standard deviation of the variable, N the sample size, π the proportion of the sample in the program, and $1 - \pi$ the proportion in the control or comparison group. To solve for N,

$$N = \frac{(\sigma / S_{\bar{X}})^2}{(\pi)(1-\pi)}$$

Thus, a larger sample size will be needed as 1) the dispersion of the dependent variable (σ) increases the confidence interval, 2) the precision $(\sigma / S_{\bar{X}})^2$ desired increases, and 3) the proportion of participants in the sample (π) moves away from 50 percent.[24]

Sources of Data on Comparison Group Members

When retrospective studies are made and require the use of existing records to identify persons to serve as members of comparison groups, the evaluator may turn to several sources of data. If he desires to identify individuals who applied for a program but who did not enter it (probably the preferred comparison group for retrospective studies), he must rely on records which usually are found, if they exist, in the local offices of the referral agency. The use of such records requires the identification of the projects to be studied, the contacting of the local offices, a detailed search of their records to ensure a complete list, and the selection of a sample from this list. Since these records probably will not contain any information about the individuals after the completion of the program, the individuals will have to be contacted or the national records checked to make certain that none of the comparison group subsequently participated in the program. These procedures will be both costly and time consuming.

---

[24] Thus, if the estimate of earnings is to be within ± $200 (i.e., within two standard errors of $100) and earnings have a standard deviation of $2,000, the sample would have to be about 1,600 (if half are participants).

This discussion and example are taken from Ashenfelter (1). For a more complex analysis, see Pitcher (22).

Alternatively, comparison groups may be constructed from among individuals similar to the participants who have been surveyed in other studies. Several studies of this type exist.

Current Population Survey. The Current Population Survey (CPS) interviews approximately 53,500 households each month and approximately 200,500 each year. Each household averages more than two individuals over age fifteen so that data are collected for more than 100,000 people each month. For each of these persons information is collected on a Control Card CPO-260, including sex, age, color, marital status, educational attainment, relation to household head, veteran status, number of family members, family income category for the preceding 12 months, occupation, industry, location, description of housing, and social security number. There is also an address and telephone number. With these data, a closely matched comparison group could be constructed for subsequent contact. If data are desired on the earnings and income of these individuals and their work experience during the previous calendar year, the surveys conducted in February, March, and April of each year gather this information. Approximately 50,000 households are asked both earnings and work experience questions, which allows for the integration of the two sets of data.

Two problems are involved in the use of these data. The first concerns access to the respondents. To eliminate the persons who have participated in the program under study, the social security numbers of the possible CPS survey comparison group members would have to be compared with any national listings for the program, which technically would not be difficult if the social security numbers can be obtained. The Census Bureau which conducts the CPS, however, must protect the identity of the persons it interviews. It must also maintain the integrity of the fundamental purpose for the CPS—to collect data on labor force behavior. The Census Bureau would have to agree either to include questions on its regular CPS surveys or to conduct special surveys after the households left the CPS interviewing cycle. Such agreement appears unlikely given the legal and other constraints on the Census Bureau.

The second problem involves the question of sample size. It might not be possible to secure a large comparison group where matching is desired on many characteristics. It would also be difficult to construct large samples to match groups who comprise a small portion of the population. This problem takes on major proportions where matching on the basis of the individual's work experience is needed (as it will be for most social programs since they deal primarily with the unemployed) and where the CPS earnings and employment data are also to be used for dependent variables. The work experience questions are asked in February and April and the income questions are asked in March. Data on these variables for the preceding year (year t - 1), are available for approximately 50,000 households. If, however, the dependent variable is data for the following year (year t), data can only be secured for the 25,000 households who are in the CPS sample for the second time (in year t + 1). Furthermore, subgroups may be difficult to find since persons unemployed at some time during a year constitute only a small proportion of the non-institutional population 16 years old or over (20 percent in 1975, which had the highest postwar unemployment). The existence of such problems can be found by comparing the size of particular subgroups in the CPS with the minimum number of comparison group members required. Because of its large size, however, the CPS should provide the necessary comparison group in most cases when fine breakdowns are not necessary.

Social Security Administration CWHS Data. The Social Security Administration maintains a special file called the Continuous Work History Sample (CWHS), providing a 1 percent sample of individuals who have applied for social security numbers. Information available for these persons includes age, sex, race, covered earnings, and employer industry and location for each quarter with covered employment. The use of these data to form comparison groups involves several problems. First, the CWHS tapes do not contain identification of individuals, although they permit linkage with tapes from the Social Security Administration through common case numbers. Some form of accommodation would have to be worked out with the Social Security Administration if social security numbers are to be compared with lists of program participants to make sure that the individuals in the potential comparison group have not been program participants too. Also, since the CWHS contains only a limited set of information, matching on these variables may not be sufficient to select a truly comparable group. Important variables such as education, marital status, health, existence of other forms of training, and family income are absent. Moreover, it would be impossible to secure this information directly since the individual's address is not a part of the record. Finally, the only dependent variables which could be examined using this source of a comparison group are the information on earnings, number of quarters of covered employment, and industry and location of employer. While the information on earnings is a key measure of the success of employment-related programs, the shortcomings of the Social Security CWHS data would appear to limit severely its usefulness as a source of comparison groups. As was discussed above, however, social security records are the best source of earnings data for long term follow-ups of comparison groups selected from other sources.

National Longitudinal Surveys. Four groups of five thousand individuals have been surveyed regularly since 1966-68 in a program sponsored by the Employment and Training Administration of the U.S. Department of Labor. The four groups are males between 14 and 24 when first interviewed, females in the same age group, women initially between 30 and 44, and men 45 to 59 at the first interview. The samples are nationally representative, and there is a three to one oversampling of nonwhites. In addition, a cohort of 13,000 youth between the ages of 14 and 21 began in 1979. This sample includes overrepresentation of blacks, Hispanics, and the poor.

The surveys include a wide range of information about each respondent, including age, sex, race, marital status, number of dependents, family income, education and training, work experience, earnings and income of the individual and spouse during the preceding 12 months, current labor force status, health, assets, family background, mobility, and psychological measures. In addition, these data are provided on a longitudinal basis so that substantial prior information on the comparison group members would be available. Finally, the data are readily accessible and extremely well documented.[25]

There are also problems in using these data to select comparison groups. They can only be used for the sex and age groups that they include. Even for these groups there may not be enough sample members in small categories of the population. Next, if the dependent variables are to be taken from the survey, the periods of work experience and earnings for the program participants must be the same as for the comparison group. The surveys have not been conducted in each year. Furthermore, the data in the longitudinal surveys are collected in specific months and the program participants must also be interviewed in these months to be strictly comparable. Also, the social security numbers of the potential comparison group members would have to be checked against the national lists to remove those persons who had participated in the programs. Finally, it is not clear whether these cohorts will continue beyond the early 1980s.

Due to the problems involved in the use of each of these sources of comparison group data, we recommend that they be used only when there is no alternative or as a supplement to other comparison groups. Limitations on the ability to match them to participants, in the methods of data collection, and in the type of information they contain create many threats to internal validity and make their exclusive use hazardous.

---

25. Data tapes may be purchased from the Center for Human Resource Research, The Ohio State University, at cost.

The Choice of Independent Variables for Analysis

To conduct an evaluation of social programs, it is necessary to measure the relationships between the program impacts (the dependent variables) and a variety of independent variables, including the personal characteristics of participants, the program components, and the conditions under which the programs operate.

Personal Independent Variables. These independent variables are particularly relevant for comparing program participants with the members of the control or comparison group to discover whether differences exist between the groups which may affect the program impacts. It is necessary to include in the analysis as many variables as possible which are correlated with both program participation and the dependent variables. However, most relevant dependent variables with which evaluations of employment-related programs deal are functions of more than one independent variable. To omit some variables in the analysis may lead to distorted conclusions due to correlation or interaction among these variables and those independent variables which are included in the analysis.[26] The analyses should treat all of the independent variables simultaneously.

The use of simple cross tabulations to isolate such relationships is inadequate in most cases. For instance, the effects of race, age, education, and skill level on earnings are all interrelated. Yet each of these effects should be distinguished. To cross tabulate by all of these variables would involve so many cells that the sample would have to be enormous. In addition, the tables would be so large as to be unmanageable. Therefore, multivariate techniques should be used in the evaluations to discover and test the statistical significance of any relationships which are observed. Multiple regression and correlation techniques can be performed with a much smaller sample than cross tabulations and permit easy interpretation of the findings.

---

[26] For a brief discussion of this problem, see Suits (33).

Evaluations should examine the effects of the programs on groups of participants for other reasons too. The analyses should determine whether or not a particular program will benefit certain target groups for whom the programs are designed, as well as find which programs serve the groups best. For most employment-related programs, independent variables should be included in the analysis to represent different groups with high percentages among the poor and the unemployed, such as blacks, Mexican-Americans, Indians, the handicapped, teenagers, the aged, school dropouts, ex-convicts, welfare and unemployment insurance recipients, and veterans. The degree of success should be measured for such characteristics of the program participants as age, sex, race, ethnic group, number of dependents, family size, education, etc. Many of the relevant personal characteristics are listed in Table 3-1.

It is also necessary to treat personal characteristics in the evaluation in order to improve the efficiency of the programs. Programs will have varying results for different types of people. The personal variables in Table 3-1 can also be used to determine which individuals get the greatest benefits from each program, and individuals can be assigned so that the success of each program is maximized. The attitudinal variables in Table 3-1 may be particularly useful for these purposes.

This assignment process may be in conflict with the desire to benefit certain target groups, however, because these groups receive lower benefits than do other workers from all of the programs. In this case, knowledge of which programs serve which groups best will still be useful because it still will be more efficient to allocate the target groups to those programs where they receive the largest benefits. If there are still program slots, the individuals who would have the greatest expected benefits would be enrolled.

Table 3-1
Personal Variables Affecting Labor Market Success

CHARACTERISTICS

age
health and disability status
race and ethnic background
sex
marital status

family composition including:
  number of other family members
  age and sex of other family members
  number of dependents

other family income and number of other family members working
educational level and quality
military service
assets and debts
employment status at enrollment

previous occupations
previous pay levels and earnings
reason left last job
previous training
skills and abilities

licenses and certifications
knowledge of job seeking skills
knowledge of how to keep jobs
socioeconomic background
jobs and connections of friends and relatives

intelligence level
arrest record
income maintenance status at enrollment
eligibility for different types of income maintenance
tax rate on earned income

attitudes of other family members toward work
access to transportation
appearance
ability to express self orally and in writing
prior mobility

ATTITUDES

attitude toward working
self-esteem
self-confidence
general disposition
motivation to get ahead
degree of independence
level of maturity of youth
attitude toward accepting welfare
willingness to relocate
occupational and pay aspirations and expectations
perceived limitations on ability

Program Component Independent Variables. Most social programs consist of a set of activities, and many of these are common to several programs. It would be extremely useful in modifying existing programs and in the planning of new programs to know which of the components are most effective for various types of participants. It would also be desirable to have information on the best combinations of components. To the extent that the length and nature of the components supplied to individuals differ within or between the programs, multivariate techniques can be used to identify effective components. If evaluations examine programs which include a variety of components and where the length of the components vary they should include as independent variables the amount of each service performed in a program (this will usually be expressed in terms of hours spent per participant) and, if possible, a measure of quality.

Exogenous Independent Variables. Employment-related social programs also differ in their effectiveness depending on the characteristics of the location and the circumstances in which they operate. Among potential factors affecting program success are: level of unemployment, growth in employment, average earnings, and the degree of manufacturing in the area in which the program occurs. The size and nature (farm, rural, depressed, etc.) of the area in which the program occurs and the degree of discrimination in the area might also be included. The type of skill, the demand for workers with the given skill, and the average earnings of persons with that skill would be important if the program involves training for specific skills. These variables should be included in the analysis as independent variables to determine under what conditions the programs are most effective and which programs are most effective under particular conditions.

Determining Proper Program Size—Measuring at the Margin

A basic question which the evaluation should answer is "What should be the program's size?" (including the possibility that the answer may be that no program is justified). Ideally, the evaluation would provide an accounting of the total benefits derived from the program at each possible level of program activity. The decision maker could then compare programs and allocate his expenditures to yield the level of activity for each which would maximize the total return on the total expenditure. To do this, he would allocate his resources so that each additional dollar was spent on the program which yielded the greatest return for that dollar, given the distribution of previous expenditures.

To date, however, evaluations of employment-related programs have not presented these data. Rather, average benefits have been calculated for a program at fixed levels of program activity. In order to make program size decisions, users of these analyses have had to assume that the average benefits of different programs have a direct relationship to the benefits at the margin, i.e., that adding a person to a program with a high average benefit would be more beneficial than enrolling the person in a program with a lower average benefit. Only if this assumption is true, however, will the decision maker end up with the optimal allocation of his resources.

There are techniques which can be used to roughly approximate the effects of changes in program size. One technique is to relate the absolute and relative sizes of the program in different labor markets to the level of program success in those labor markets. There is a wide range of program sizes which may be used to predict the effects of program growth or cutbacks. For example, if the program has higher average benefits in areas where only 5 percent of the poor participate as compared with areas with higher participation rates, program expansion would be expected to reduce average benefits, all other factors held constant.

As changes occur in program size, this usually means changes in the type of program participants. For instance, small programs may "cream" and only recruit high school graduates while larger programs may have to dig deeper and enroll dropouts. It is likely that programs have differential effects depending on the type of participants. Therefore, if the evaluations can determine the average effects of the programs on different groups of participants, this can be used to predict the effects of changes in program size with increased participation by particular groups.[27] Similarly, if changes in program size involve changes in program components, knowledge of the average benefits for each component will be useful. As the number of evaluations increase, more exact measures of the effects of program size will gradually become available. Presumably the experience of older programs can then be transferred to new ones. Meanwhile, however, we should seek to measure the impact of programs on subgroups of the participants, for components of the program, and in different areas.

---

[27] The average benefits of a group may not equal the benefits for the last member of that group who participates in the program. If, however, the effects of the programs on the groups under consideration differ greatly, the use of average benefits for subgroups will probably approximate more closely the marginal effects of the change in program size than will the average return for all program participants.

Exercise 3-1

Podunk Community College (PCC), which serves Podunk City but not its suburbs, is eager to expand its services to the community. One of the areas it has identified to serve is the unemployed and underemployed older professionals in the community. A course has begun at PCC on how to write a resume, how to prepare oneself for an interview, and how to utilize proper techniques for job search. The first class in the course was enthusiastic about it, and the course was favorably reported in the local newspaper. In that news story, the president of PCC was quoted as saying, "It is our duty to serve all of the people of Podunk. We will provide this course to any resident of Podunk City who wishes to enroll. Further, no one will have to wait to enroll. We will run as many sections of the course as are necessary to accommodate all potential students."

Dr. M. A. Lee, the Director of the State Education Department, reading the news story, questions if this course should be offered at other community colleges in the state. She asks you to evaluate the success of the course as it has operated at PCC. Before you actually begin the study you are required to present her with an outline which: (1) shows the research design you will use to measure the changes caused by the course, (2) presents alternative designs, and (3) presents arguments for your suggested design and against the alternatives. Prepare such an outline.

Exercise 3-2

Write a proposal which responds to the following Request for Proposal (RFP).

SPECIFICATIONS FOR MEASURING THE EFFECTIVENESS OF COUNSELING IN THE UNITED STATES EMPLOYMENT SERVICE[28]

I. *Overall Purpose of the Research*

In recent years, questions about the value and effectiveness of the U.S. Employment Service (ES) counseling program have been raised. As part of a series of ES-related research projects, the Department of Labor (DOL) plans to examine the extent to which counseling, as formally delineated within and made available through the facilities of public ES offices, makes a difference in the subsequent labor market experiences of those counseled. Such an examination, however, involves complex issues of measurement and methodology, requiring careful exploration before any full-scale evaluation can be undertaken.

Therefore, the primary purpose of this project is to develop and pretest an experimental design for use in measuring the independent net effects of counseling and in assessing the overall effectiveness of the ES counseling program. This developmental work is expected to provide initial information, for consideration by ES administrators, about the independent net outcomes associated with ES "exemplary" counseling programs, i.e., counseling programs established and maintained in accordance with objective standards of quality.

If a feasible experimental design is developed, a separate and comprehensive evaluation project may then be attempted. It would be competitively procured following the completion of this proposed developmental research.

---

[28] This case was excerpted from a Request for Proposal issued by the Manpower Administration, U.S. Department of Labor on May 16, 1975 as RFP 7[illegible]3.

II. *Background and Description of the ES Counseling Program*

The importance of employment counseling in manpower programs has been recognized by Congress through the passage of more than 22 pieces of legislation specifically calling for this service. The relevant legislation includes the Wagner-Peyser Act of 1933, the Manpower Development and Training Act of 1962, the Economic Opportunity Act of 1964, the Vietnam-Era Veterans Readjustment Assistance Act of 1972, and more recently, the Comprehensive Employment and Training Act of 1973.

The U.S. Employment Service and its affiliated state agencies have responded to these various pieces of legislation by establishing and maintaining a nationwide counseling program which now employs approximately 3,000 counselors at an estimated annual cost of over $20 million. Some of these counselors work only part time in the program, however, and the size, scope, and quality of the program varies from local office to local office.

The primary role of a counselor in the ES is one of assisting applicants who have problems of occupational choice, occupational change, or occupational adjustment which prevent them from being satisfactorily referred to a job. An employment counselor is necessarily concerned with the outcome of job placement, whether in terms of direct placement through the ES, or by the facilitation of job placement at the highest level consistent with an individual applicant's skills, experience, and aptitudes. An employment counselor, however, is also concerned with a variety of other outcomes which contribute directly or indirectly to employment.

In fiscal year 1974, approximately 980,000 or 7.4 percent of the 13.3 million persons who filed applications with local ES offices received counseling services. However, an informal examination of ES applicant records has suggested that, in any recent year, roughly 20 to 25 percent of ES applicants have been in "need" of counseling services (as "need" is defined in the ES manual).

III. *Objectives, Scope, and Limitations of the Research*

For purposes of developing and pretesting the experimental design to be used in a comprehensive evaluation of the ES counseling program, and of providing initial information on the extent to which "exemplary" counseling programs make a difference in the labor market experiences of those counseled, the research must focus on the following objectives:

(1) an analysis of the findings, i.e., of the differences in outcomes between those applicants who *needed and received* counseling services (the experimental group) and those who *needed but did not receive* such services (the comparison group);

(2) the specification of an experimental design for use in a comprehensive evaluation of ES counseling; and

(3) a discussion of critical methodological and/or procedural problems encountered during the course of the developmental research, including recommendations for overcoming such problems in a subsequent evaluation.

IV. *Methodology of the Research and Technical Requirements for the Proposal*

Bidders are expected to propose and defend their methodologies for carrying out the research, generally within the framework provided by this RFP. Proposals should address all questions or issues which are important to meeting the objectives of the study. The choice of methodology, the identification of outcomes to be measured, the intervening variables to be considered, as well as the rationale used to justify these choices will be evaluated in terms of how well they demonstrate an understanding of the complexities of the research as they relate to ES operations.

The research may be carried out on either a concurrent or a retrospective basis. Any retrospective study should include ES applicants only from the 12-month period preceding the collection of data on outcomes. This should enhance the researcher's ability to locate subjects for follow-up purposes. Because of certain drawbacks associated with each of these methods, however, the bidder is expected to discuss and defend the rationale for and the implications of his preferred approach, in contrast to the alternative. Optional approaches, carefully considered and argued, may be proposed.

The purpose of having a developmental phase of work is to see whether a methodology for measuring the independent effects of counseling can be developed. Thus the offices selected for examination should meet some set of standards which can be used to define an adequate level of counseling program effort. If it is not possible to devise a means of detecting differences in outcomes between the experimental and control groups within "exemplary" offices (i.e., those meeting certain standards used to define counseling quality), it would make little sense to undertake a full-scale evaluation whose sample would include offices which fail to meet those basic standards.

There are, at the present time, approximately 2,400 local ES offices throughout the country. Of these, about 600 have staffs of 20 or more and employ at least one full time counselor. The sample of "exemplary" offices to be included in the research should be drawn from a universe of approximately 150 offices which have counseling programs that meet specified standards of adequacy. The DOL will review the appropriateness of the actual standards for sample selection proposed by the bidder.

The experimental groups should be drawn from local office applicants "needing" (as defined in the ES manual) and receiving counseling services. The comparison groups should be drawn from local office applicants "needing" but not receiving counseling services.

As a summary checklist, proposals are expected, as a minimum, to discuss the following methodological issues:

(1) whether the research will be carried out on a concurrent or retrospective basis;

(2) procedures for selecting the offices to be sampled and criteria for selecting the applicant groups for inclusion in the samples;

(3) problems one should expect to encounter in trying to establish comparability between and among the experimental and control/comparison groups, and the methods that will be used to deal with such problems; and

(4) the sources of data to be collected, and the timing of data collection.

Chapter 4

THE COSTS OF SOCIAL PROGRAMS

No program provides its benefits free; something must be given up in order to derive them. Resources which are devoted to social programs cannot be used to produce other goods and services. Society, by devoting human resources to conduct training programs for painters, loses the services of the instructors as painters, as auto assemblers or even as college professors, to use a few examples of some of the alternative production which may be forsaken. Thus, the costs of social programs should be considered to be their opportunity costs—the value of the alternative benefits which are foregone because of the programs.

Estimating Opportunity Costs

It is usually impossible, however, to identify what benefits are foregone and to place a value on them, especially for large social programs. For instance, what production would be lost if society chose to require all students to have an additional year of high school? What potential output did society give up by providing medical services to the poor and old under Medicaid and Medicare? Consequently, because of the measurement problems the opportunity cost is often assumed to equal (or at least be proportionate to) the market price of the resources which go into the program. This assumption rests on competitive market theory which states that in a perfectly competitive market, each marginal factor of production will be paid an amount equal to the value of its product in its present and next best alternative use. Thus, in perfect competition the amount paid for producing a product will reflect the value of the opportunities foregone. (This means that instructors in the painting program example will be paid an amount equal to the value of their output in their next best occupation, say, as painters.)

There are many problems with the application of this theory. Social programs do not usually operate in situations of perfect competition. Most social programs involve governments purchasing resources which do not have large, well-defined competitive markets with many alternative buyers. For example, the number of purchasers of educational buildings and personnel is quite limited in most geographic areas—there are usually only a few school districts and proprietary schools. In addition, the government is not a profit-maximizer. (One of the many assumptions of perfect competition is profit-maximization by all parties.) Therefore, the government may pay economic rents to the owners of resources, i.e., the government pays more than is necessary to attract those resources from alternative uses. One such example might be the hiring of unemployed workers for public service employment. Even though the alternative uses of these workers may be very limited, the government usually pays at least the minimum wage and possibly a considerably higher one. The social program expenditure in this case can include a form of transfer payment as well as an expenditure to cover the opportunity cost.

Other deviations from the perfect competition model include monopoly power, externalities, and nonmarginal purchases. In some cases the sellers of resources may possess monopoly power which permits them to obtain economic rents from those operating social programs. An example might be the imposition of regulations requiring the staff of a program to be residents of the area in which the program is operated. This may severely limit the number of eligible persons and allow them to gain salaries above their opportunity costs.

In the case of externalities, the amount paid to the resource will not reflect the true opportunity cost for society because of unmeasured or unattributable costs. An example of such an externality is the situation where participants in a social program forego employment in order to attend. There may be persons who would otherwise be unemployed or employed at lower paying positions that are able to fill the jobs left vacant by the participants. Society in this case loses the output of the program participants but gains the output of the secondary groups. Overall there may be no net loss to society.[1] Alternatively, if the program involves work experience or on-the-job training, a program participant may be displacing someone else who would have been doing this work. If this second displaced person becomes unemployed (or causes someone else to become unemployed), the output of the program participant should not be included in its entirety as a negative cost in calculating the costs of the program to society.[2]

Finally, there are cases where the purchases made for social programs are so large that they affect the price of the resources being purchased. In these nonmarginal cases the problem is whether the opportunity cost of the resources is equal to their old price, the new price, or a price between the two.

Even with these problems, however, the opportunity costs are assumed to be reflected by the costs of the resources involved. This occurs because no alternative method of valuing foregone opportunities offers a better solution. It is necessary to be cognizant of the possible shortcomings of this approach, however, and to make adjustments when appropriate. (The adjustment is often referred to as "shadow-pricing"—making an estimate of what the price would have been had a competitive market existed for the sale of the resource.) Also, luckily, the problems are not as severe with employment-related social programs as in other areas. Most resources are bought in relatively open markets.[3]

---

[1] Similarly, the government will not necessarily lose tax revenues or pay increased transfer payments even though the participants pay less and receive more from the government. When the secondary individuals find employment, the government welfare and unemployment payments to them may fall to the extent that they could offset the government losses on the participants.

[2] Again, calculations of government costs should take the displacement effect into consideration by estimating the reductions in tax receipts and the increase in transfer payments which result for the second party.

Listing of Costs of Employment-Related Programs

Obviously, it is important to distinguish whose benefits are being foregone. We may have the situation where the party who is doing the giving up is not the same as the one who is receiving the benefits. One of the most common mistakes of program evaluation is to ignore the costs incurred by parties other than the government.

Society, individual program participants, employers, and government may each be required to give up resources for use in the programs. In some cases, expenditures of resources will mean foregone opportunities for more than one group. For example, salaries of program administrators will be costs for society as well as for the government. There will also be resources foregone which will be costs for some groups but which will be gains for other groups. For example, government allowance payments to program participants or reimbursements paid to employers will be costs for the government but will actually reduce the costs of the participants and the firms involved.[4] Therefore, we once again present separate lists for each party.

A. COSTS FOR SOCIETY

Society's costs for operating social programs consist of the *real* resources (goods and services and not merely funds) which are used up by the program but would have been devoted to other productive uses had the program not occurred. (Transfers of goods and services, such as changes in welfare payments, within the society are not counted since they do not alter the amount of real resources used, only the parties who use them.) Among the resource costs of social programs are the following.

---

[3] For more detail on costs see Judy (20) and Smith (30).

[4] As a general rule, all changes in resources and funds which occur during the program are considered as costs, and all changes after the program are considered as impacts. There may be gains during the program, such as higher income for the participants, but it is easier to treat these as negative costs.

1. The Time Spent by All Personnel Involved in the Program.

   Special care should be taken to include the time of the following groups who are often not considered, particularly those who are not engaged full time in the program.

   a. *The Local Project Staff.* The costs should include the value of the time they spend on activities such as the design of the project proposal; recruitment, testing, and counseling of prospective participants; the provision of any supportive services connected with the program as legal services, counseling, custodial care of the equipment and facilities used by the program, day care services, health services, remedial education, and transportation; processing allowance payments; job development; skill training; job referral and placement; follow-up counseling; evaluative follow-up; and all of the record keeping and other administrative tasks involved in each of the programs.

   b. *Persons at the Regional and State Levels.* Included would be all persons who are in any way involved in the program. This would include consultants to local projects, field supervisors, persons responsible for project approval and review, and statisticians involved in reviewing project reports. Again, their salaries should be allocated in proportion to the time spent on the program.

   c. *Personnel in the National Offices of U.S. Departments or Other Agencies Involved in the Programs.* Their salaries should be allocated in proportion to the time spent in administering the program, including budgetary review, fiscal accounting, policy planning, project approval, project monitoring and evaluation, research, and the training of staff and technical assistants.

2. The Physical Capital Used in the Program.

   This would include the market rental value of all property and buildings, including government property, and the market rental value of all machines, instructional equipment and supplies, and other materials used in the program. Equipment which is purchased should be depreciated based on use. Where it is not possible to estimate depreciation on a use basis, the difference between original cost and salvage value should be amortized appropriately over the life of the program.

3. Miscellaneous Services.

   All services necessary to the operation of the program, such as staff travel, telephone services and equipment repairs should be measured

4. The Goods and Services Purchased by the Program Participants Which They Would Not Otherwise Have Bought.

   These include such expenditures as transportation to and from the program, the extra cost of meals and living expenses away from home, uniforms, books, tools or other educational materials, and day care for dependents.

5. The Potential Production Which is Lost.

   During the time the program is being conducted, the participants may work less than they would have worked were there no program. This loss of potential output must be considered a cost to society. Of lesser magnitude, if there are individuals involved in the operation of the program as unpaid volunteers, their potential production during the time they devote to the program should be measured.

B. PARTICIPANT COSTS

1. Out-Of-Pocket Expenditures for Items or Services Participants Would Not Have Purchased Were They Not in the Program.

These items are discussed above in section A4. They represent the participant's money which is spent on the program rather than used for other purposes, e.g., for housing, entertainment, or deposits in a savings account.

2. The Income Loss for the Participants While They Are in the Program.

This would consist of the after-tax earnings lost because of participation in the program minus any increased government payments received as an inducement to enter the program. The latter would include training allowances, subsistence allowances, travel allowances, wages, and higher welfare or unemployment insurance payments than would otherwise be received.

C. EMPLOYER COSTS

The opportunity costs for employers are the unreimbursed expenditures of company funds or property devoted to the program which could have been used to increase profits or dividends. These would include the following expenditures if they are not offset by a reimbursement from the government.

1. Wage Costs of Company Employees Who Perform Services for the Program.

2. The Value of All Physical Capital Used Up in the Program Which is Owned by the Firm.

3. Miscellaneous Expenditures by the Firm on Services Necessary for Program Operations.

   Included would be spending for materials and supplies.

4. Production Lost Because of Poor Quality of Work Which Results as a Part of the Learning Process for the Participants.

   From this cost any reduction in taxes as a result of lower profits as well as any reimbursement from the government should be subtracted.

D. TOTAL GOVERNMENT COSTS

Government costs can be viewed in two ways: a) as the total increment in expenditures and reduction in revenues accruing to a unit of government; or b) as the expenditures from a particular line in a budget, labeled for the specific program. The first views the costs from the perspective of the impact of The first views the costs from the perspective of the impact on a total government budget. The second takes the perspective of the program manager. The latter narrower definition of costs is quite simple to measure--it is the amount of funds allocated for the program which nave been spent.

The total government cost concept must consider more than the increment in expenditures and use of government resources for the specific program being examined. It must also include the increase or reduction in the use of resources for all other programs which occurs because of the existence of the program under examination. For example, a program manager of a Vocational Rehabilitation (VR) program may try to reduce his costs by enrolling illiterate VR clients into an adult basic education program sponsored by the local school system or by CETA rather than providing the services directly through VR. This would reduce the expenditure of the VR program funds but total cost to the government would be the same.

The total government cost concept would examine the following categories of costs for providing services to the participant less the cost of services which would have been provided to the participant in the absence of the program.

1. The Costs of All Personnel for Whom Government Pays the Salaries or Reimburses Others.

   Personnel to be considered include those discussed in section A1 and those personnel for whom employers are reimbursed.

2. The Value of All Physical Capital Used in the Program Which is Government Owned or Rented, or for Which Government Reimburses Someone Else.

3. Expenditures on Miscellaneous Services Provided by Government or for Which Payment is Reimbursed by Government.

4. The Net Increase in Government Payments to Individuals Made to Induce Them to Participate in the Programs.

   These would include the items discussed in section B2.

5. Tax Revenues Lost During the Program.

   These would include the reductions in personal income taxes, social security taxes and unemployment insurance taxes which may result from lower earnings and the reductions in sales and excise taxes caused by lower expenditures of the participants while they are in the program. In addition, lower corporate profit taxes might result from decreased efficiency of participants involved in on-the-job programs.

6. Other Items for Which Government Makes Payments.

   These would include payments to firms conducting on-the-job training to compensate for the lower productivity of trainees.

Measuring the Increment in Costs

Just as the proper measurement of program impact was based on the comparison of outcomes with and without the program, this should also be the key to proper measurement of program costs. One should consider as costs only those additional resources that would not have been used in the absence of the program. For example, most employment-related programs involve a job referral component. When calculating the cost of this component, the expenditures that the Employment Service would have made for job referral for these individuals if the program had not existed must be subtracted from the program expenditure on this item. Only if this subtraction is made can the increment in the resources used by the program be calculated, and it is only this concept which can be compared with the measures of additional well-being which were discussed earlier.

Thus, to measure many of the costs requires knowledge of what would have happened to the program participants had they not participated. The best way to get this knowledge is to use a control group which is randomly selected from persons willing and able to enter the program. Only this group will give an internally valid estimate. Thus if the costs are to be as accurately estimated as program outcomes, the same type of control group must be used to measure both. If projects are selected for evaluation when funded, this will permit the same control groups to be used to measure the costs and impacts of a program.

The Use of Control Groups. Control groups should be used to provide information for three types of cost estimates. The first is the loss incurred while the participants are in the program. While participating in the program, individuals usually are not engaged in what they normally would otherwise be doing. Therefore, participation in the program may lead to losses of after tax earnings, unemployment compensation, or welfare payments by the individuals; potential production by society; and taxes by government. The experience of the control group during the course, however, should not be affected by the program. (This assumes that the program is not large enough to affect the entire labor market.) Therefore, the difference between the control group's and participants' after tax earnings, unemployment compensation, welfare payments, production, and taxes will show the losses actually incurred because of participation in the program.

The second use of control groups is to determine how much of the governmental services received by the participants would not have been received if there had been no program. Earlier we discussed Employment Service job referral services which are normally used by many of the persons who enter employment-related programs. Similarly, when welfare recipients enter such programs the counseling they receive in the program may merely replace counseling they would have received from a case worker. Therefore, it is important that information be collected on the amount and nature of all governmental services received by both the participants and the control group. If this is known, the latter can be subtracted from the former to find the actual increment in services which result from a program. Then, only the cost of this increment in services should be compared with the benefits which were calculated as the differences between the two groups.[5]

Finally, the control group can be used to measure the increment in program-related expenditures by the participants. Some programs require that the participants incur expenses for travel, instructional materials, uniforms, living expenses and meals away from home, etc. Some of these expenditures represent added costs of program participation; others may not, however. If an individual would have been taking the bus to work instead of taking it to a training center were he not in a program, there may be no additional cost of transportation resulting from the program. To arrive at this conclusion it is necessary to know the expenditures associated with the program by the participants and the expenditures on these items by the control group.

If it is not possible to have a control group, the same information should be gathered for a comparison group. All of the same internal validity problems would occur in the measurement of costs as were discussed with reference to benefits in Chapter 3. Therefore, once again we strongly recommend that control groups be used.

Measuring Costs for Other Family Members. In addition to the control group, measurements should also be made for other persons who might be affected by the program. One group which is very likely to be affected is the participant's family. For instance, if participation lowers the earnings of the participant, the slack may be taken up by another family member who accepts a temporary job that he would not normally have taken. Such changes could be discovered by comparing the work experiences during the program not only of the participants and members of the control groups, but also of their respective families.

---

[5] The implicit assumption in the use of a control group to measure the increment in services is that the level of service provided by the program being evaluated does not alter the amount of service provided by other programs. In terms of the examples, the funding of additional Vocational Rehabilitation programs should not reduce the amount of the Employment Service or public assistance case worker aid provided to the control group.

Accurately Measuring the Costs

Government accounting systems are designed on an appropriations basis and not on the basis of incremental costs. Typical government cost accounting says that anything paid out in the name of a program is a cost of that program. As a result, many items are improperly charged to a program while other costs are ignored. Some specific examples of these problems follow.

Proper Program Assignment. The costs of functions performed for a program sometimes are not charged to that program. We have already discussed the example where one operator of a program may devote staff time to job development, while another may use Employment Service (ES) staff to carry out the same function. Under present accounting procedures, these costs are all assigned to the program in the former case and all assigned to the ES in the latter. Yet the same services are being performed for the program participants in the two areas. Other examples of costs that are often improperly assigned are remedial education, which is sometimes provided by the local school systems; the use of school buildings and equipment for training programs; the time spent by various public officials preparing project proposals; and the value of the services of persons on loan from business to the government or of volunteers.[6] Therefore, in these and many other situations one must go beyond the costs directly assigned to a program in order to include the costs of all of the additional services provided because individuals participate in a program.

Similarly, if expenditures are assigned to one program but in fact are made in part for other programs, only a portion of their costs should be included. For instance, equipment purchased for one program subsequently may be used for others (such as machines bought for a training class which are subsequently used in other classes or for vocational education purposes), or persons who are hired for one project may also be used on other projects.

---

[6] Valuing the opportunity cost of volunteers' time is especially difficult because the volunteers may be using leisure time. A truly satisfactory method for assigning a dollar value to leisure has yet to be developed. Presumably, it is at least as valuable as the income the individual could earn were he employed. (If leisure is less valuable, he would go to work.)

Inclusion of All Costs. All of the costs of a program should be measured. Yet, there are several items frequently overlooked. The first, the value of public facilities provided free of charge was discussed above. While the use of these facilities does not represent an outlay of funds, there are still costs involved. Most equipment will wear out with use, and its use by social programs will accelerate the need for replacement or repair. Therefore, the depreciation of this equipment should be calculated and included in the costs of the programs. This will be particularly important for the often quite expensive instructional equipment used in vocational training courses.

Even for buildings that will suffer little from the additional wear, there are still costs associated with their use by social programs. Other activities may be displaced. For instance, if a vocational high school is used for CETA training, vocational education classes may have to be held elsewhere or may not be held at all. Both of these alternatives will involve costs. Similarly, even when an abandoned army base is used as a Job Corps camp, there are costs if the property could have been sold or leased to private industry. Thus, to measure the true cost of these facilities, they should be valued at their market rental value.

The administrative cost of social programs is another category typically underreported. There is a large administrative overhead connected with each program. Many of these items are included in the outline above. The time that civil servants at the national and regional levels spent on planning the original program, budget and proposal approval, project review and monitoring, program-connected evaluation and research, fiscal appropriations and accounting, and on all other administrative duties involved in program operation at these levels should be considered. Ideally, the time spent by all government workers from the Secretaries of Health, Education and Welfare and Labor on down, should be apportioned among programs and among particular projects, if possible. Greater attempts should be made to approximate the services provided in each of these categories since they represent sizeable costs which presently are excluded from most calculations.

Calculating Marginal Costs

Another important concept for measuring both impacts and costs is that they should be measured at the margin. Social programs are evaluated for the most part in order to say whether they should be expanded or contracted rather than whether they should be maintained at their present size or eliminated. Therefore, what is needed are the benefits and costs associated with various changes in their size. For instance, what will be the difference in total benefits and costs of adding an additional person to a class or an additional 10,000 participants to a program? Often the cost of adding extra participants is considerably below the average cost for an existing program. For example, adding one person to a lecture-style class is almost costless, in most cases, since the classroom and instructor are already committed. The result is a declining cost curve such that the marginal cost is considerably below the average cost and the benefit-cost ratio at the margin is much higher than it is on the average (assuming constant benefits). Yet, generally only overall costs and benefits are considered and not those at the margin. Only if the ratio of average benefits and costs is proportionate to benefits and costs at the margin for all programs will average calculations of benefits and costs be accurate guides for making program decisions. The problem, however, is how to measure benefits and costs at the margin.

One way of looking at the problem is in terms of the total national program. If it has had a long history and has operated at a number of different sizes there may be a sufficient total number of points to construct a total cost curve; the first derivative of that curve will yield the marginal cost of the program. One can similarly calculate marginal benefits. This procedure requires a number of different studies over a broad range of program sizes and the real world operation of social programs does not lend itself to this type of analysis because most of the programs have recently been developed or modified.

Another method of calculating marginal cost is based upon how much it costs to add additional individuals to particular projects.

In this situation, projects are the unit of observation. Arraying projects by size and cost per project will show the marginal cost of adding individuals to projects. This can be done more formally by regressing project cost on project size to calculate the marginal cost as was done by Somers and Stromsdorfer (32). This procedure, however, has a potential limitation stemming from the possible correlation between the size of the program and the labor market conditions in which it exists. Thus, a program of 500 in New York City may yield costs that are quite different from a program of 500 in a small town.

An extension of the regression technique was used by Hardin and Borus (17) when they regressed total cost on the size of the project in terms of total enrollees, the number of sections of the course, the number of weeks it ran, and the number of hours per week. This analysis was designed to show not only the effects of increasing the number of enrollees but also the possible effects of changing other components of the magnitude of the program. In this analysis it was found that it was somewhat cheaper to increase the number of enrollees per section than to lengthen the course or expand the number of sections, at least in the instructional cost area.

Joint Costs

Joint costs exist when the use of resources produces more than one type of output. The classic example is the raising of sheep to produce both mutton and wool. In this situation, once the sheep have been raised for wool there is no additional cost to raising them for mutton. Or alternatively, once raised for mutton there is no additional cost to having them produce wool. The question is "When a resource is to be used for one of these purposes, should any cost be assigned to the other purpose?" For example, machinery bought for use in a vocational education class is often used in a CETA training program. Some people have argued that since the primary use of the equipment is in the vocational education programs and it has been purchased for that purpose, there is no need to charge any costs to the training program since the marginal cost of that item is zero. While such logic is true for the sheep's wool and mutton, it is not true for the type of items being discussed in most social programs. This is demonstrated by identifying the opportunity costs involved. The use of the machine in vocational education courses does not prevent it from having an opportunity cost in alternative uses at other times. Presumably, if the machine was not being operated during the training program, it would be available to someone else who might rent it. Similarly, public buildings used for social programs have separate opportunity costs for each period that they are used so their rental value in these uses needs to be calculated. There appear to be very few cases where social programs actually do use existing facilities which do not have some alternative use, and hence a cost which is indistinguishable from the social program use. Social programs are, fortunately, not like sheep.

Examining the Costs of the First Program Participants

The cost of a particular program will vary, depending on when in the course of the program's development the measurement is made. As with most businesses, the costs of social programs will decline as their scale and life increase. A program will have relatively high costs per participant when it is being developed and the number of participants is small. Program staff will be engaged in training personnel, establishing bookkeeping systems, writing proposals, experimenting with program ideas, and similar organizing functions. As the program becomes more established and experience is gained, less time will be spent on these activities and more time can be devoted to providing services, increasing the number of participants, and reducing unit costs. While the benefits may not change with program size,[7] it is almost certain that the program costs will change. Therefore, the time period in which the costs of a program are measured will be crucial, particularly at the beginning of a program.

---

[7] It is possible that benefit levels will also change as the program develops. This could occur if, as the program operators devote more time to providing services, the quality of such services as training improves.

To resolve the problem for a new program, we suggest that the benefits of the first program participants should be related to the costs for later participants in the program.[8] While this would cause the costs assigned to the first group of program participants to be below those actually incurred, it would give a better indication of the long run costs of the program after it has become more established. Moreover, it would not lead to any delay in evaluation since, as has been mentioned, the success of the participants should not be measured until one year after they leave the program.

---

[8] Cost data still should be gathered at the start of the program to make comparisons with later cost figures, to find the extent of decline in costs with increases in the size of the program, and for administrative purposes.

Exercise 4-1

The following table lists the costs of a vocational education program run jointly by a school system and employers. Under this program students study half of the day in the school and then learn a trade in an employer's establishment. At the top of each column are the parties who might incur the costs. Place a plus (+) in the box if positive costs are incurred, a minus (-) if negative costs are incurred (i.e., if the party receives resources instead of losing them), and a zero (0) if there are no costs to the party.

	Parties Paying Costs			
Type of Expense	**Society**	**Government**	**Employers**	**Students**
I. Vocational School Expenses A. Personnel—salary, fringe benefits and value of payment in kind made to: 1. Teachers, group leaders, and other types of instructors				
2. Administrators (principal, secretaries, etc.)				
3. Other institution personnel (janitors, tool crib keepers, food workers, etc.)				
4. Administrators outside the institution (planners, administrators of the program at the regional and national levels)				

B. Operating Expenses 1. Office supplies consumed during school year				
2. Travel by staff				
3. Transportation of students				
4. Building maintenance and equipment repairs (general upkeep exclusive of personnel)				
5. Supplies provided by the school to students (books, metal, wood, etc., used by the students in their work at the school)				
6. Miscellaneous operating expenses				
C. Capital Expenses 1. Buildings (rental value of the building)				
2. Land (rental value of land)				
3. Equipment (rental value of equipment)				
II. Earnings Foregone A. Expected earnings of the students were they not in school during school year				
B. Earnings of the students during school				
C. Value of the services of volunteers who aid teachers were they to be purchased				

(continued)

Type of Expense	Parties Paying Costs			
	Society	Government	Employers	Students
D. Taxes which are not paid because of reduced earnings (1) by students,				
(2) by volunteers				
III. Student Out-of-Pocket Expenses A. School supplies (tools, uniforms, etc.)				
B. Transportation to school and work				
C. Expenses which would have been paid were students not in school (work supplies, transportation)				
D. Tuition, if charged				
E. Transfer payments received by the students (welfare payments, living allowances, tuition reductions, other allowances which they would not receive were they not in school)				
IV. Employer Expenses A. Supervisors' salaries and fringe benefits for the time they are supervising or instructing the students				
B. Supplies and materials used as part of the teaching process				
C. Wages and fringe benefits paid to the students				

D. Other costs of having students such as overhead for space and equipment				
E. Value of the students' output on the job				
F. Payments received as compensation for the training (1) from the government				
(2) from students				
(3) other sources				

Exercise 4-2

The Podunk Vocational Rehabilitation office has encountered a problem. Clients have not been keeping appointments for weekly counseling sessions or have been showing up late. This has meant that the staff have periods when they are not being used while they are overloaded at other times.

Dr. Judy, the office director, calls a staff meeting to discuss the problem. She explains that having counselors idle is both inefficient and costly. She proposes that clients be scheduled for appointments only at 8:00 A.M. and 1:00 P.M. When they come in they will each receive a number and wait until their number is called. The system will be first come-first served so that those who do not want to wait will come in promptly when the office opens at 8:00 A.M. or reopens after lunch at 1:00 P.M.

Ms. Mickey, one of the counselors, argues that such a system is not desirable because it is unfair to the clients and the office reception area will be crowded and noisy. Instead, she says the office should copy the airlines and "overbook." Counselors can normally see two clients an hour. To reduce unused time for "no shows," three clients should be scheduled for each hour's time. If all three clients appear, requiring 90 minutes instead of 60 minutes, those scheduled for later appointments will have to wait. Dr. Judy asked what would happen to clients still waiting at 5:00 P.M. when the office closed. Would they be rescheduled for another day? Ms. Mickey thought they should be seen that day with the counselor receiving overtime pay. This was vetoed by Dr. Judy because it would break the office budget. Mr. Fish, representing the clients, argued that the whole discussion was making a mountain out of a molehill. He noted the building, lights, heat, and staff would all be there and paid for whether there were ten or one hundred clients served on a particular day. Therefore, it was no more costly to keep the present system than either of the proposed solutions.

What are the costs of each of the three alternatives? Explain which one you would choose and why you would select it.

Chapter 5

COMBINING THE MEASURES OF PROGRAM IMPACT AND COST

The data gathered for evaluations of social programs should provide the information to answer four types of questions. 1) Should a particular program be continued? 2) Which of several alternative programs should be expanded or contracted? 3) In what ways can changes in the components of a particular program lead to improved efficiency? 4) What programs best serve particular groups of individuals? The data discussed in the previous chapters will provide the answers to these questions. Before such answers can be obtained, however, several technical decisions must be made. These include: how to handle potential impacts which occur after measurement, how to treat impacts and costs which occur over time, and how to handle external effects and secondary impacts.

Treating Future Impacts[1]

Social programs may affect the participants for the remainder of their lives. Therefore, to accurately assess all of the impacts of these programs it would be desirable to wait until all persons in the program have died.[2] Obviously such a suggestion is ludicrous. The purpose of program evaluation is to make policy decisions on whether the program should be continued, and if so, in what form and for whom. The decision maker cannot wait thirty to seventy years before he makes a judgment. We have suggested that the measurement of impacts should occur one year after the program's conclusion, with subsequent follow-ups. This will require estimating what impacts will occur after these measurements are made. Such estimation requires two sets of projections: how will program impacts change over time and how long will the impacts last?

---

[1] We concentrate on questions relating to impacts since we have defined all effects, positive or negative, which occur after the program as impacts and those which occur during the program as costs. Most of the impacts after the program should involve positive benefits.

[2] Since there may also be intergenerational effects, an even longer period of observation may be needed.

How the Impacts Might Occur Over Time. It is conceivable that the impacts of an employment-related program might grow, remain constant, or decline as time passes. They might grow to the extent that the individual participating in the program is placed on an entirely new and higher job ladder that he would not have attained in the absence of the program. A good example here would be the vocational rehabilitation program which provides artificial limbs. Most of this program's clients would have little expected earnings during the rest of their lifetimes in the absence of extensive prosthetic treatment, physical therapy, and vocational training. With the training and the prosthetic devices they are able to function, at least partially, in the employment world and to receive the earnings increases most people usually receive as they build seniority.

Alternatively, one can see situations where the effect of a social program might be to shift the individual up one rung on a ladder above where he would have been in the absence of the program but move him no further. For instance, an upgrading program within a plant might take a semi-skilled machine operator and train him to be a machinist. Once he achieved the machinist status, however, his percentage increase in wages would not be very different from those paid to a machine operator.[3]

---

[3] Even here, however, the individual will have increasing benefits over time in an absolute sense because the percentage wage increases are calculated on a higher base; that is, a 5 percent increase on a $2.00/hour base would give an extra ten cents an hour, whereas the same 5 percent raise for someone earning $3.00/hour would yield fifteen cents an hour.

Finally, there may be a situation where the impacts of the program decline as time passes. This may be due either to obsolescence of the skill and training provided or to the fact that other avenues of advancement would have opened to the program participants had the program not existed. Examples of the former case are a number of employment-related programs which have provided training for jobs with specific employers or training in the use of particular types of equipment. The person who has been trained for the needs of a specific employer may find when he leaves that employer his specific skills are of no use to other firms hiring in the job market. Similarly, the skills of a person trained on a particular piece of machinery may become at least partially obsolete if that piece of machinery or the production process changes.

An example of a situation where program participants may be given a jump ahead of individuals who do not participate (but these other individuals may subsequently find other means for catching up) would include the individual who receives vocational training in high school. He may have a temporary advantage until others subsequently receive such training in a postsecondary school, in the armed forces, or in on-the-job training. Thus, his initial advantage may be eliminated with the passage of time and the use by others of alternative methods for advancement.

Given these three possible divergent streams of impacts over time, which one is appropriate? The most logical method for making projections would be to base them on the experience of participants in other programs. If the gains from a similar program have increased at an annual rate of 5 percent, then this figure could be applied. Unfortunately, there is no reliable basis for comparison. Most social programs and almost all evaluations of them are less than fifteen years old. Longitudinal data are not yet available on them. Further, the few longer term studies which have been conducted with comparison groups are contradictory. Hu *et al* (19) and Somers and McKechnie (31) found declining earnings impacts for vocational education after six years and for institutional skills training after five years; Ashenfelter (2) found declining earnings gains for males but relatively constant gains for females over a five-year period following MDTA institutional training; Borus (5) found increasing gains for five years following institutional training; and Borus and Prescott(7) found increasing earnings benefits for men completing institutional training but declining gains for the dropouts.

The Duration of Impacts. If it is found that the gains of program participation decline steadily over time, they may eventually reach zero.[4] Alternatively, if they remain constant or increase during the period of observation, they may continue until the individual retires or dies. The magnitude of the projected impacts and the resulting program evaluation will be greatly affected by which one of these scenarios is selected and by how far into the future the projections are made. Different studies have chosen one or the other alternative.

A Possible Solution. The best method for estimating future impacts appears to us to be a sensitivity analysis which projects the impacts for several periods with the impacts increasing, remaining constant, and declining.[5] A matrix can be constructed which presents the expected impacts under each of the alternatives (e.g., Table 5-1). Such a matrix will show the sensitivity of the impact estimates to various combinations of assumptions, which should be useful when comparing different programs.

As longitudinal studies provide more hard data on the trends in impacts over time, the matrix can be condensed. Until then, considerable thought should be given to determining which description of the particular impact studied is most appropriate. Filling in the matrix is only the first step in projecting impacts—theoretical or empirical considerations for selecting the "best estimate" must follow.

---

4. All of the longer evaluations have found some benefits to persist for five years after the treatment.

5. Longer term projections also should take mortality and changes in labor force participation rates into account.

Table 5-1. Total Impacts Under Alternative Assumptions

Period of impact life	Impacts decline annually by:		Impacts are constant	Impacts increase annually by:	
	5%	15%		5%	15%
5 years					
10 years					
20 years					
Until participant reaches 65 years					

Assigning Values to Future Impacts and Costs

The benefits of different types of social programs may be realized over different periods of time. Likewise, various impacts of a program occur at different points in the time following participation. Finally, program impacts occur in the future (after the program), whereas costs are incurred in the present (during the program). The basic question to be answered is "Are impacts to be treated equally regardless of when they occur; if they are not, what is the proper method for equating future impacts with those which occur now or costs which have already been incurred?"

To construct a simple example, take three programs that require an investment in 1980 of $80 million. Program A will return $80 million in 1981 with no further impact. Program B will return $40 million of benefits in 1981, $40 million of benefits in 1982, and $40 million of benefits in 1983. Thereafter, there will be no benefits. Program C will provide benefits of $20 million for each of the seven years from 1981 through 1987. There will be no benefits after 1987. Which program is preferable or should there be no program?

Program A is unlikely to be the choice because most people are not indifferent toward options giving them an amount of income in the future equal to the amount that they invest now. This occurs for two related reasons. The first of these is "time preference." This is the old saying, "A bird in the hand is worth two in the bush." Simple proof of this desire to have our cake now rather than in the future is that most of us require that we be paid a certain amount in order for us to postpone consumption. In order for the bank to induce us to make deposits, they must make interest payments. Similarly, most people borrow money and pay interest in order to buy things now rather than to wait until they have the necessary cash for the purchase. Individuals will pay 12 percent interest charges in order to have a new car now rather than saving the money and buying it two years from now. The individual's rate of time preference, which is subjective, can be estimated by his willingness to loan or to borrow. He will loan money as long as the return is greater than his time preference rate and will borrow money as long as the interest rate he has to pay is less than his time preference rate.[6]

Another way of looking at the problem of valuing future benefits in terms of present expenditures is in terms of the opportunity cost of alternative projects. If a person can earn interest of 6 percent by putting his money into a riskless savings account, he would be foolish to invest in a project which paid him less than 6 percent. Similar logic applies to government projects. The government should not undertake any project which pays less than the alternative investments that it might undertake or the return the resources would earn in the private sector.[7] The use of private sector rates of return is important because taxation of the private sector is the source of the resources which will be used to undertake the project.

The two concepts are interrelated. In a perfectly competitive capital market, as borrowers whose time preference is high borrow more, the interest rate will rise and their time preference will fall. Similarly, the interest rates lenders can charge will fall and their time preference will rise as they choose the best of alternative projects. Eventually, an equilibrium point will be reached where the borrowing rate, the loan rate, the rate of time preference, and the opportunity cost of alternative projects will be equal. This will be "the" discount rate—the rate by which future returns would have to be reduced to find their present value. The formula for determining the present value of impacts is:

---

[6] It must be recognized that there are factors other than time preference—such as precautionary saving for a rainy day—which will influence willingness to borrow and loan at various interest rates.

[7] This statement holds only if the returns from the projects are fully and accurately measured. It may be that the government project has certain noneconomic benefits which are socially desirable and which cannot be expressed in economic terms. In this case, the government project might be undertaken even though it does not yield the rate of return attainable in the private market.

$$PV_B = \frac{B_1}{(1+r)} + \frac{B_2}{(1+r)^2} + \ldots \frac{B_n}{(1+r)^n} = \sum_{t=1}^{n} \frac{B_t}{(1+r)^t}$$

Where B_1, B_2, B_n are the benefits (net of losses) in years 1, 2, and n, respectively, r is the rate of discount, and t is the year. In terms of our example, if the discount rate is 10 percent, the present value of the benefits of the three programs are $72.7 million for Program A, $99.5 million for Program B, and $97.4 million for Program C.

Unfortunately, in the real world the capital market is not perfect. There are several major factors which prevent the establishment of a single discount rate. One is the tax rate on private projects. Since the corporate profit tax rate approaches 50 percent, a project needs to have a rate of return at least twice as high as the individual's rate of time preference in order for the project to be worthwhile for him to undertake. Another consideration would be that projects are not equally risky. An individual, in making his calculation of the rate of return, must include some premium for the risk involved in the project. Therefore, the rate of return may have to exceed the rate of time preference before it will be worthwhile for the individual to undertake a project. Furthermore, the rate of risk will vary among projects so that there will be a whole variety of interest rates charged to arrive at an equally riskless rate of return. Other disrupting factors which lead to a diversion of interest rates from the single discount rate are various market imperfections, such as lack of communication, imperfect knowledge, monopolies, and outright discrimination. Finally, there is a problem of externalities in that the rate of return in the private market may not take into account all of the costs or the benefits which are derived from the project. This is a problem of proper allocation of costs and returns to a project. For instance, the owners of a polluting smokestack may not be charged for the pollution. The pollution exists, however, and it lowers the rate of return on the project for society although it does not lower the private rate of return. In conclusion, there is not a single discount rate which can automatically be applied to all social projects.

Which rate, then, should be used? As stated above, the discount rate should be no lower than the amount people are willing to pay to borrow funds to increase consumption, nor should it be less than the return the funds could earn in alternative investments. These rates differ, however, depending on who is borrowing and who is investing. The government borrowing rate, practically risk free to the lender, has recently been between 4 and 9 percent. Business faces a somewhat higher rate—the prime rate has been between 6 and 12 percent in recent years. Individuals face many borrowing rates ranging from about 8 percent to over 25 percent, depending on the risk involved to the lender. The alternative investment opportunities also vary. The basic alternative to a government program is a tax cut which will allow increases in private business investment. The return on private investment is in the range of 10 to 20 percent, part of which is a premium for risk. The return to the individual is typically lower (5 to 15 percent), again dependent on the risk involved. Since the extent of risk is not known in the investment cases, the true rate of return is not known for alternative projects.

To quote Prest and Turvey (23) "the truth of the matter is that, whatever one does, one is trying to unscramble an omelette, and no one has yet invented a uniquely superior way of doing this." Therefore, the most logical way to proceed is to consider a variety of possible discount rates and then to test how sensitive the analysis is to each choice. Rates of 5 percent, 10 percent, and 15 percent would appear to represent a reasonable set. They should cover the range of time preference rates. For most social program comparisons the relative values of the impacts of two programs will vary little with the choice of discount rate. In some situations, however, substantially different results will occur.[8]

If only a single discount rate is to be used, it would appear that the consensus is now at a rate of 10 percent, particularly for discounting society's future benefits. This rate roughly approximates the return on private investment when a small allowance is made for risk. However, other rates should be used as well in order to demonstrate the sensitivity of the analysis to this assumption.

Accounting for Externalities

Up to this point, governmental and societal impacts and costs basically have been calculated by aggregating the impacts and costs associated with the program participants. We must now be concerned with external impacts and costs—benefits and costs accruing to persons who are not directly participating in the social program.

The question may be raised as to why these outside individuals should be considered. The answer is that these individuals may be affected in ways that completely offset any value of the program. For example, it is conceivable that program participants are placed in jobs which would have gone to other individuals had the program not existed. Further, it is possible that the program participants replaced individuals with identical characteristics so that there is no net increase in employment in the aggregate or for any particular group in society. Alternatively, the program participants may replace other individuals who hold political power with the result that the program becomes politically infeasible because the latter refuse to allow the program to exist. Similar issues revolve around the question of who pays the costs of social programs.

---

[8] In terms of our example, a 5 percent discount rate will cause Program C to be preferable, whereas Program B has the highest yield using a rate of 10 percent. If an 80 percent rate is used, Program A is to be the best of the three alternatives but no program would be preferred to all three.

There are two types of external benefits and costs. The first are real externalities. They involve the creation of additional real production (real goods and services which would not have existed before), or they use additional real resources (and therefore they reduce the amounts of goods and services which can be produced). The real external effects are particularly important when one is discussing the impacts and costs of social programs from society's point of view.

There are several simple examples of real external effects of employment-related social programs. First, there are the effects on other members of the participant's family. For instance, an individual may be provided with services which increase his earnings. This in turn raises the family's income. With the increase in family income, the spouse and/or some other family member may reduce the number of hours they work. Whereas the family may be happier with this allocation of labor market effort, the increment in production is only the change in the output of the total family unit, and this is less than the increase in the production by the program participant. The loss in production which results from the reduced work effort of other family members must be subtracted from the increase for the participant.

Another example of effects which are external to the individual participant but are internal to his family would be the influence of the program on the investments in education and human capital made by other family members. For instance, the higher earnings which may result from a social program may be spent on the program participant's children's education so that their productivity is subsequently increased. There would be intergenerational effects of the programs which increase society's output of goods and services, although over a relatively long period of time in this case.

A third example of real externalities occurs once the program participants are employed. If they interact favorably with other workers in their place of employment, the program participants may either transfer skills to the other workers and/or form a work group which operates very well together. The result may be that the other workers receive higher earnings due to increased production as a result of the existence of the employment-related program.

In all three of the above cases, there are obvious effects on the real output of society produced by workers other than the participants in the social program. There is another class of externalities called pecuniary externalities. In these cases, there is a redistribution of income to individuals external to the program but there is no change in the real production of these individuals. For example, if a large number of workers are trained for a particular occupation, it is likely that the wage rate will decline (or not increase as rapidly) for all workers in that occupation, including those who had been formerly employed there. Since there is a change in earnings but no change in the output of these other workers, this is a pecuniary externality and not a real one. Society has not lost any production from these workers and, therefore, societal impact of the program is unaffected.[9] It should be pointed out, however, that this redistribution of income from the pecuniary externality may be extremely important to the individuals who have suffered from it; thus, they may attempt to thwart the continuation of the program.[10]

There are four other external effects which are very important for employment-related social programs—the displacement, vacuum, substitution and complementary effects. They may be either real or pecuniary externalities, depending on the circumstances.

Displacement Effects. The displacement effect occurs when a program participant either causes a presently employed worker to lose his employment in a given firm or, what is more likely, causes some individual who would have been employed in that firm not to get a job because it was taken by the program participant. The most overt example of this is an employer who hires program participants and keeps them for as long as he can use them at some subsidized rate, but then replaces them with other workers who are subsidized. This of course is a rare case. A more common occurrence is the situation where the program participants possess somewhat greater skills due to the program than do other workers who are available in the labor force. The program participants will be hired rather than the other workers because the firm presumably will reap greater profits by hiring these more productive individuals.[11]

---

[9] The government impacts also may not be changed. Although tax revenues will decline from the other workers, this should be offset by higher taxes paid by other factors of production.

[10] One can argue that the reluctance of many groups to accept equal employment opportunity programs is an attempt on their part to avoid the negative pecuniary externalities associated with these actions. A prime example is the building trade unions who refused to admit large numbers of journeymen or, for many years, to cooperate with government apprenticeship programs for blacks.

Whether displacement causes real or pecuniary externalities depends on how well the labor market operates, particularly how closely the labor market approximates full employment equilibrium and the time period that is involved in movement from one equilibrium to the next. If the assumptions are made that there is a full-employment equilibrium at the start of the program (all workers are being used to their fullest capacities) and that there is rapid adjustment to a new full-employment equilibrium, then those workers who were displaced would be able to move into alternative occupations which at the margin pay the same wage rate as they were formerly earning. Any deviation in the increment in output for society from the increment in the earnings of the original program participants will depend on the speed of adjustment to the new equilibrium. In the case where the adjustment is instantaneous, displacement does not lead to real or pecuniary externalities (assuming the adjustments are small enough so that the wage rates of the displaced workers will be unaffected).

---

[11] The extent of displacement will depend upon the shapes of the demand and supply curves. In the usual case where the demand curve is downward sloping and the supply curve is rising to the right, there will be less than total displacement. The degree of displacement will rise as the elasticity of the demand curve declines and the elasticity of the supply curve rises. If the demand curve is vertical or the supply curve is horizontal, there will be total displacement; if the demand curve is horizontal or the supply curve is vertical, there will be no displacement.

If, however, there is less than full employment so that there are losses involved in moving from one equilibrium to the next, then there may be real externalities involved in displacement. For instance, if workers experience unemployment as they are bumped down from one occupation to the next, this period of unemployment involves a real cost to society because these workers are not producing goods and services. Moreover, to the extent that some of the displaced workers do not find employment, their lack of employment must be considered as a negative impact of the program. The crucial question in examining displacement effects is how long it takes for the adjustment process to occur and to what extent displaced workers are likely to be unemployed. It would appear that the higher the rate of unemployment, the longer will be the period during which the displaced workers are unemployed and the larger the real externalities.

Vacuum Effects. The vacuum effect is in some respects the reverse of the displacement effect. If individuals who participate in social programs would have held jobs had they not been in the program, the effect of program participation is to reduce the supply of labor in those occupations which they left. This might increase the wage rate in the occupation and attract workers from other occupations with the result that the reduction in employment in this occupation would be less than the number of persons who left it to enter the program.[12] Again, whether there are real or pecuniary externalities will be a function of labor market conditions. There are likely to be real externalities if there is less than full employment which allows some workers to take jobs when they formerly would have been unemployed or underemployed. This means that the opportunity cost to society in terms of lost production will be less than the output not produced by the participants while they are in the program. In contrast, the vacuum effect may be negligible or have no impact at all if there is full employment and the number of workers involved is small.

---

[12] The size of the vacuum effect will depend on the Elasticities of the demand and supply curves. There will be a partial effect if both curves slope in the normal directions, no effiect if the demand curve is vertical or the supply curve horizontal, and a total effect if the demand curve is horizontal or the supply curve is vertical.

Substitution Effects. There may also be displacement of workers in industries other than those in which the program participants are placed. Most employment-related programs do not create a demand for the services of participants. Rather, they must assume a given distribution of a fixed level of demand, which dictates job openings.[13] Under this assumption, however, an increase in the supply of labor in a certain occupation would lower the wage rate, increase the supply of goods to meet orders which could not previously be filled because of a shortage of labor, or cause displacement. The first two possibilities would increase employment in the occupation and the volume of goods produced by that occupation. If, however, the aggregate level of demand is fixed, there could be reductions in the demand for the goods produced by workers in other occupations as there is the substitution of program participants' products for the other workers' goods and services.[14] This would mean a reduction in the earnings and possibly in the employment of persons in these other occupations.[15]

In the case of the substitution effects, the nature of the externalities again depends on the extent and ability of the market to move between full employment equilibria. If there are, for instance, restrictions on the downward movement of wages in industries where a decline in demand for the product causes labor to fall, then some of these workers will become unemployed. If they can find other comparable work easily, the losses will be minimal in terms of output foregone. However, there will be substantial real external effects if they remain unemployed for long periods.

---

[13] Presumably, government fiscal and monetary policy are designed to achieve a given level of employment regardless of whether a particular social program is implemented. At the same time, however, programs which increase the employability of the labor force and thus reduce manpower bottlenecks will allow full employment at lower levels of price increase. This consideration would be of special importance if monetary and fiscal policy were restricted to less than full employment levels because of inflationary problems.

[14] The extent of substitution depends on the cross elasticity of demand curves for the products produced by the program participants.

[15] For example, if I want to purchase orange juice but the price is too high or I will have to wait too long for it to arrive, I buy milk instead. When, however, the price of orange juice falls or it becomes more generally available because of a training program to increase the number of orange pickers, I will shift my demand and buy the orange juice. This shift in demand allows the employment of more persons and greater income for the orange industry. It hurts the dairy farmers and processors, however; it lowers their revenues and possibly their production. If we measure the effect of the program solely in terms of the orange industry we may be missing some of its effects.

Complementary Effects. Finally, there may be real positive externalities and social gains even if there is no unemployment, when the placement of the program participant also leads to others being hired. Technology is such that jobs are complementary in some industries. If one job is vacant, other jobs which are dependent upon it cannot be filled. For example, a shortage of computer programmers will limit the use of computers and the need for keypunch operators. In these cases, the improvement in the earnings and employment of the program participant (who becomes a computer operator in the example) understates the benefits which society receives. The complementary workers (keypunch operators) have higher earnings or more employment than they would have; these, too, are benefits resulting from the program. The extent of such complementary effects will be based on the labor market situation and technology in the industries in which program participants are placed. For the complementary effects to occur there need to be bottlenecks, which are most likely to occur only as the economy approaches full employment. It would appear that if the participants fill job openings by displacing other workers there are likely to be no complementary effects.

Measuring the Externalities. By definition, externalities are difficult to measure. They require identification of the parties outside the program and the comparison of their situations in the presence of the program with what would have happened in its absence. For members of the participants' families, such procedures are relatively easy—the families of control or comparison group members will represent the situation without the program.

The other externalities do not lend themselves to these measurements. There is no way to identify positively the individuals who would have been hired were the program participants not hired first, nor is there a way to identify the individuals who are hired when program participants do not apply for a job that they would have filled were they not in a program. The best that appears possible is to realize that the externalities are dependent on the level of the economy and the size of the program, to examine them carefully, and to present some alternative assumptions about the displacement, vacuum, substitution, and complementary effects.

Secondary Impacts. In the preceding section the discussion concerned direct effects of social programs on persons external to the programs. There may also be indirect effects of the spending for the programs—the secondary or multiplier effects of the programs. In its simplest form, the logic runs as follows. Social programs represent an increment in expenditures. As the monies for the program are spent by the program participants and the program staff, the result will be greater income for other individuals who receive the funds. These individuals will, in turn, spend more. To the extent that resources are unemployed, the expenditures will cause increases in production through the multiplier effect. These secondary effects could also be included among the impacts of the program.

Most evaluations, however, ultimately concern the choice of alternative program expenditures or a reduction in taxes. These alternatives will also have secondary effects, and it becomes merely a question of how the multiplier will be called into effect and not whether it will be applicable. Since all programs will have a multiplier effect, there seems no point in considering it directly. One need not multiply the benefits by a fixed factor (such as a multiplier of two or three) since all programs will presumably have such a multiplier.[16]

---

[16] An argument can be made that the size of the multiplier will vary depending on the type of expenditure and how it is used, e.g., training programs might have a higher multiplier than public service employment because in the latter case some of the expenditure may replace local funds. Also, government expenditures may have a higher multiplier than a tax cut. Estimating the magnitudes of the differences, however, is not practical and, therefore, they are probably best ignored.

While the use of an overall multiplier applied across the board does not appear useful, the geographic distribution of programs may cause them to have different impacts because the degree of unemployment varies considerably between areas. There will be pockets of unemployment and depressed regions even in periods of relatively full employment nationally. Spending in areas where there is high unemployment will produce greater real production gains than will equivalent amounts spent in areas of full employment, where the additional spending will only lead to higher prices. In the case of employment-related social programs, the effects are probably much more localized than the effects of alternative development projects. Most of the expenditures for the program would go toward the purchase of services and income maintenance payments rather than for capital goods. Since the expenditures for services and income maintenance would be spent in the area in which the program occurs, most of the program expenditures remain in the community, at least in the first round. Of course, as the income is spent on manufactured goods in the second round, it will most likely move outside the geographic area in which the program occurs. One would expect to find, however, greater increases in real productivity in areas with high unemployment, than in those where there is full employment.

While it is possible to indicate the direction of secondary effects, as was the case for externalities, it is impossible to measure these effects with certainty. We, therefore, are faced with the situation where we must attempt to estimate the external and secondary effects of social programs whenever possible while, at the same time, realizing that we are probably able to touch only the tip of the iceberg. It is because of our ignorance of the submerged part of the iceberg that we should be very cautious in using the results of our analysis.

Making Program Decisions

In Chapter 2 we outlined a series of many different types of possible criteria for measuring program impacts. Any or all of these criteria may be considered to be important by a decision maker evaluating a program. Therefore, a measure of program impact and cost should be calculated for every criterion which has been examined in the analysis.

Combinations of Program Impact and Cost Measures. The basic tools for combining measures of program impact and cost are benefit-cost and cost-effectiveness ratios.[17] These express the total or average amount of success (net of losses) per dollar of cost. Further, since the benefits and costs occur at different times, the ratio should show the present value of each, i.e.,

$$\frac{\text{Present value of benefits}}{\text{Present value of costs}} = \frac{\sum_{t=1}^{n} (B_t/(1+r)^t)}{\sum_{t=1}^{n} (C_t/(1+r)^t)}$$

where B_t and C_t are the benefits and costs for each year in which they occur, r is the discount rate, and t is the year. Usually, both terms in the ratio will be positive. This indicates that while the program produces benefits it involves the use of resources to accomplish this gain. When a benefit-cost ratio is greater than 1, it indicates that the present value of the economic returns exceeds the present value of the costs of the program. If the ratio is less than 1, this indicates that the program costs more than the value of the resources gained from it. For noneconomic gains, a subjective weighting is necessary.

Whereas the benefit-cost ratio is probably the most widely used criterion for evaluating projects, it suffers from several possible shortcomings. First, if either term in the ratio is negative there are problems of interpretation. A negative numerator and a positive denominator indicate that not only are the original costs never recovered, but in addition, further losses are incurred after the conclusion of the program. On the other hand, if the numerator is positive and the denominator is negative, the program not only generates successful outcomes on its completion but also provides more resources during the program than it consumes. If both the numerator and denominator are negative, the net gains are generated during the period of the program while net losses occur after the completion of the program. Such a situation is highly unusual.

---

[17] The expression is a benefit-cost ratio when the criteria of success are measured in dollars (such as increases in participant earnings) so that the numerator and the denominator are expressed in dollar terms. If the numerator is expressed in a unit other than dollars (such as number of persons employed or change in score on a job satisfaction scale), the expression is a cost-effectiveness ratio.

A related problem with benefit-cost ratios is that one must be clear as to what constitutes a benefit and what constitutes a cost. Earlier, we suggested that costs include those changes which occur during the course of the program and that impacts are any changes which occur after the end of the program. However, because different programs occur over different periods, it becomes somewhat difficult to apply the definitions of benefits and costs equally to all programs. Yet, this can become crucial in the calculations of benefit-cost ratios. For example, some employment-related programs provide on-the-job training or work experience where the participants produce useful output while they are in the program. The magnitude of the benefit-cost ratios will differ, depending on whether this output is considered as a benefit and added to the numerator or a negative cost and subtracted from the denominator.

A third problem with the benefit-cost ratio is that it requires that a rate of discount be established to find the present value of the benefits and costs. As discussed earlier, the selection of a particular discount rate may be subject to question, particularly since the choice of rate may determine the magnitude of the benefit-cost ratio. However, one does have the opportunity to make a sensitivity analysis to find out how the benefit-cost ratios for different projects vary in their ranking if different discount rates are used. When such differences occur, another combination of benefits and costs may be more appropriate.

When the numerator or the denominator of a benefit-cost ratio is negative or there are substantial problems in differentiating benefits and costs, another combination of impact and cost measures—the net present value which gives the net value of the gain from the program—is usually more appropriate. This may be expressed as:

the present value of benefits minus the present value of costs

or

$$\sum_{t=1}^{n} (B_t/(1+r)^t) - \sum_{t=1}^{n} (C_t/(1+r)^t)$$

It shows the present value of the additional resources which have been gained after costs have been deducted. Therefore, the net present value must exceed 0 in order for the program to cover its costs.

While the net present value combination solves the problems of negative costs or benefits and the definition of benefits and costs, it, too, suffers shortcomings. These include the problem of choosing a discount rate. In addition, it should be obvious that the net present value can only be calculated when both the benefits and costs are expressed in the same units, as is also true for benefit-cost ratios. One cannot subtract apples from oranges or dollars from units gained on a satisfaction scale. Also, when comparing alternative programs, the net present value is appropriate only when the expenditures on the alternatives can be the same. Otherwise, there will be a tendency to choose the larger project. For example, using present values for both cases, Project A may have a cost of $10 million and a benefit of $100 million while Project B has a cost and a benefit of $100 and $200 million, respectively. The net present value of Project A will be $90 million while that of Project B is $100 million. Project B appears superior. Yet, if Project A could be expanded it would be much more profitable.[18]

---

[18] To mitigate this problem it may be useful to calculate the net present value per dollar of cost. Once this is done, however, the problems of definition of benefits and costs and the treatment of negative benefits and costs, as observed with the benefit-cost ratio, reappear.

A third measure, when both the numerator and denominator are expressed in dollars, is the internal rate of return. This is the annual discount rate which will equate the total benefits and costs. To estimate the rate of return the following equation is solved for r:

$$\sum_{t=1}^{n} \left(\frac{B_t}{(1+r)^t} - \frac{C_t}{(1+r)^t} \right) = 0$$

For a program to be successful, it must have a return greater than 0. To be better than an alternative program, it must have a higher rate than does the alternative.

The obvious advantage of the internal rate of return approach over the others is that it does not require an explicit discount rate and thus avoids the problem of choosing one. Also, it makes no difference whether a particular change is labeled as a negative benefit or a positive cost. The internal rate of return suffers, however, from other shortcomings. Most important of these is that it cannot be calculated without the use of computer facilities.

Should a Program be Continued? The answer to this question usually depends on what alternative programs are available. General agreement should exist on the discontinuation of certain types of programs. A program should be ended when no redeeming features are found after consideration of all criteria of success—where all important dependent variables are measured and: 1) no benefit-cost ratio is greater than 1; and 2) no cost-effectiveness ratio has a positive numerator or a negative denominator. These criteria will be met very infrequently if only because it is usually impossible to quantify all of the dependent variables. Therefore, the program decisions must be based on a comparison of alternative programs.

Comparison of Alternative Programs. In very few cases, one program is superior to another when compared on all of the criteria that we have suggested. In these cases the course of action is clear; the superior program should be expanded. (This is based on the assumption that average benefits and costs are positively related to those at the margin.) More often, one finds a program which is superior in some areas but inferior in others. The choice of program expansion and contraction under these circumstances depends upon the preferences attached to each of the benefits. For example, a skill training program may be more effective than a remedial education program in raising the earnings and reducing the unemployment of the participants, but the remedial education course may lead to greater personal satisfaction and improvement in race relations.[19] Assuming that only one program can be expanded, a choice must be made as to which is more important—increased earning and employment or psychic and behavioral improvements. Once explicit weights showing relative importance are assigned to each of these impacts, the program decisions can be made. The weights should be explicit so that others who have different values can also use the analysis.

There are two strategies which may be followed in assigning relative weights to program impacts. The first is for the decision maker to provide the evaluator with the weights of various benefits before the evaluation is begun. The evaluator will then examine only those measures of success with non-zero weights and will aggregate his findings to arrive at a single overall measure of program effectiveness. The advantage of this approach is that it does not consider goals deemed irrelevant (those given no weight by the decision maker); therefore, it may be more economical and efficient. Its major shortcoming is that the weights assigned to impacts differ among decision makers and over time. For this reason the alternative approach usually is more practical.

---

[19] So far the discussion has concentrated on the economic benefits, where the numerator can be expressed in dollars. Although noneconomic benefits cannot be stated in comparable terms, it is extremely important that these benefits not be ignored. It is useful, therefore, to present the noneconomic cost-effectiveness ratios in tabular form for the decision maker to use along with the economic data. These will take the form of ratios of the increment in a noneconomic benefit divided by the cost of achieving that increment. For instance, the remedial education program may increase personal satisfaction by 2 points on a scale of 20 at a cost of $50 to achieve that increment in satisfaction. The ratio then would be 1:25. The skill training might increase the personal satisfaction by 1 point for a cost of $100 for a ratio of 1:100. It will then be up to the decision maker to assign relative weights to the benefit-cost ratio and to each of the cost-effectiveness ratios.

In the second strategy the evaluator calculates the benefit-cost or cost-effectiveness ratio for every impact which might be relevant for each program being examined. If consideration of all possible impacts is not possible because of cost or other limitations, the calculations should at least be made for all impacts which are thought highly relevant. The ratios for alternative programs can then be compared in a single table.[20] This procedure allows each decision maker to assign the weights that he believes are most appropriate and to arrive at a decision of overall program value. If circumstances change, the decision maker can redefine the weights he wishes to use and simply recalculate the relative performance of the programs. The weights should be determined independently of the results of the analysis; otherwise, there is a great post-analysis temptation to find the weights which will make the analytical results conform to previous prejudices.

Improving Program Efficiency. The same procedures could be used to compare the benefits and costs of the components of a particular program in order to determine the most efficient combinations of components. The multivariate analysis proposed earlier includes the effects of the presence, the duration, and the quality of program components on each of the measures of success and cost.

There may be a component which has no benefit-cost ratio greater than 1 and no positive cost-effectiveness ratio for any of the possible criteria. Such a component probably should be dropped. In some cases, however, components have to be treated as complementary sets. For example, diagnostic testing by itself

---

[20] Such a table might take the following form:

	Benefit-Cost or Cost-Effectiveness Ratio		
Criteria of success	Program A	Program B	Program C
Increased production	3.0	2.0	1.0
Reduced unemployment	1 week/$1,000	2 weeks/$1,000	3 weeks/$1,000
Increased taxes	2.0	2.0	2.0

will make no improvement [illegible] the individual's behavior; without it useful counseling may be extremely difficult. It is more likely, however, that components will vary in their effectiveness, depending on the criterion of success. Once more, a tabular listing for each component can be made, including the benefit-cost or cost-effectiveness ratio for each of the criteria in order to facilitate the choice between components.

Matching Clients and Programs. Finally, the same method of analysis and presentation may be used to identify the effects of various programs and components on different types of participants to determine the best combinations for particular groups of potential clients. The multivariate analysis would show whether programs or components produce differential success or costs, depending on the types of participants. From these data, benefit-cost and cost-effectiveness ratios for a particular group of participants could be calculated for all programs and components. Once more, the weighting of tabularly presented values will allow cross-program and cross-component comparisons.

Steps for Evaluating the Impact of Employment-Related Programs: A Summary

In our opinion, the impact of all employment-related social programs should be systematically evaluated with such evaluations beginning as soon as each program becomes operational and continuing on a periodic basis. The preceding discussion provides what may appear to be a confusing number of alternative approaches to measure the impacts and costs of employment-related programs. Here we wish to compress that discussion and to outline what we feel would be a useful procedure for measuring program impact. The steps include the following:

1. Examine the components, clients and operating conditions of the program.
    a) Determine what is the nature of the program, how large it is, and with what other programs it is comparable.

   b) Identify who are the program's clients and how they are recruited and selected.

   c) Determine in which labor markets the program operates, who are the program operators, and how the program is supported.

2. Determine the possible impacts of the program.

   a) Determine the goals of the program as perceived by the commissioners of the evaluation.

   b) Determine the goals of "significant others" who may make use of the evaluation—program operators, politicians who originate or review such programs, other evaluators (possibly through publications), and program participants.

   c) Review evaluations of similar programs to determine the impacts they used and especially those which were significant.

   d) Think of all other possible impacts.

   e) Identify the impacts with the party being affected—society, government (or a particular unit of government), employers, participants, and participants' families.

3. Establish one or more measures of each impact.

   a) Use existing scales and instruments when possible.

   b) Apply marginal productivity theory to estimate society's gains from individuals' earnings gains.

   c) Be concerned with the validity of the measures.

4. Determine which impacts should be measured.

   a) Follow the principle that it is better to measure too many impacts than to omit one which is important, noting that priorities for programs change with economic conditions and politicians.

   b) Eliminate only those impacts which are impossible to measure, which would cost too much to measure, or for which no significant effect is anticipated.

5. Establish a measurement design which will have internal and external validity in estimating the net impacts of the program.

   a) Note the threats to internal and external validity.

   b) If at all possible, opt for random assignment since this provides for internal validity—make the case for random assignment as strongly as possible.

   c) If random assignment is ruled out, the next best alternative as far as internal validity is concerned is probably the regression discontinuity design. This design is less useful for generalizing.

   d) If a nonequivalent comparison group is used, note the possible threats to internal validity and realize that the superiority of one group is likely to influence the direction and magnitude of the findings. State the expected biases.

6. Determine the potential costs of the program.

   a) Avoid the use of budget line expenditures as the sole measure of costs.

   b) Apply the concept of opportunity cost to measure benefits foregone due to the existence of the program.

   c) Use marginal productivity theory to aid in the estimation of society's opportunity costs.

   d) Identify the party bearing each of the costs—society, government (or particular units of government), employers, participants, and the participants' families.

7. Apply a design which measures the net increment in costs.

   a) Comparisons of participants and control groups should yield estimates of opportunity costs for participants and production lost by society.

b) Governmental costs should include the services provided to participants and control group members to estimate the net difference.

8. Design procedures for gathering data and analysis.

   a) Determine the appropriate sample size consistent with the variation in the data and level of confidence desired.

   b) Allocate the sample selection among projects of differing types and sizes. Impact and cost data should be collected from different-sized projects, if possible, to allow measurement of marginal impacts and costs.

   c) Study the strengths and weaknesses of alternative data sources and data gathering techniques, noting that they and the sample size must conform to available funds. Wherever possible rely on an existing data source rather than establishing new data collection procedures. If the project involves contacting persons after the program, efforts should be made to keep track of the individuals so their location is known if they move.

   d) Data to be used to measure impacts should be collected for at least one year following the program to avoid seasonality.

   e) Cost data should be collected after projects have become fully operable to avoid initial high average costs due to program start-up.

   f) During the project, records should be maintained on the nature, length, and cost of all services performed for each participant.

   g) Records should also be kept for each participant indicating changes in his knowledge and skills and his attitudes toward the program. For each person who drops out of the project, an exit interview should be conducted to determine the reasons for leaving the project and the quality and nature of his job if he leaves for this reason.

h) Impact and cost data should be identified with individuals so that separate calculations of program effectiveness can be made for different groups of potential clients.

i) Use multivariate analysis to adjust for differences between the control or comparison group members and the participants. Calculate the degree of success on each criterion for the total program, for groups of participants in the program, for various program components, and for various conditions under which the program operates.

9. Adjust the impacts and costs.

a) Project the impacts for their expected duration and include any expected growth or decline in them.

b) Discount the impacts and costs to find their present value.

c) Estimate the effects of externalities—displacement, vacuum, substitution, and complementary effects—and secondary impacts.

d) Estimate the influence of threats to validity in the evaluation design.

10. Make calculations and discuss the findings.

a) Present marginal cost-effectiveness or benefit-cost measures (benefit-cost ratios, net present value, and internal rate of return) for each impact under consideration so that the reader can aggregate impacts using his own weights.

b) Aggregate impacts if agreement has been reached on a preference function.

c) Present separate calculations for significant groups of clients, for different combinations of program components, and for projects operating under differing conditions.

d) Indicate how these calculations are subject to the assumptions made in the estimation process.

e) Present your best estimate for the effectiveness of the program.

f) Suggest any improvements in program efficiency which are derived from the analysis.

Limitations on Evaluation Techniques

There are three basic limitations on the impact evaluation methods we have presented which should always be kept in mind. First, the analysis will be based only on some of the criteria of success and only on some of the costs of social programs. Our lists of impacts and costs are admittedly incomplete. Even if more were added to each list, constraints on time, funds, and ability would limit the materials which could be considered and important criteria might be omitted. More important, however, we consider only those criteria of success and costs which can be measured; ignored are those which cannot be measured. For these reasons the judgments which are made about the programs, components, and participants may not always be "right."

Second, even for measurable criteria, correct evaluations of social programs cannot always be designed. We have argued that random assignment to experimental and control groups will permit considerably more accurate and trustworthy evaluations than will other methods of sample selection. There are still threats to external validity involved, however, which may limit the generalizability of the findings. Not all variables which could conceivably influence the outcomes and costs of social programs can be included in any analysis. Therefore, an observed relationship may be due to some other factor which is correlated with the two variables being examined. For example, it may be found that on-the-job training has higher benefit-cost ratios than does institutional training. If, however, the local unemployment rate is not a variable in the analysis, and the analyzed courses are in labor shortage areas, the return to on-the-job training might fall if that program replaced institutional courses operating in depressed areas. Likewise, we measure relationships which exist for a given period of time. They may change over time as other unanalyzed variables change. Thus, while the ratios obtained from this type of analysis may be accurate for the group under study, they may be difficult to generalize.

The third limitation has been discussed before. The estimates of benefits and costs are usually averages for the programs. They do not measure directly the effect of changes in program size. While one program may have considerably higher average ratios than another, increasing the size of that program may lead to smaller gains in success per dollar of cost than would occur by increasing the program with the lower average ratios. Until further information is available on marginal gains and costs, we can only assume that the effects of changing the size of programs will be directly related to their ratios of success to costs.

Finally, there is a problem which this primer hopes to solve. Evaluations to date have been on a program by program basis, each using different methodologies. Typically, each study uses a different type of control or comparison group and a different way of handling externalities and secondary effects. Data sources, types of questions, and measures of benefits will differ with each study. Different periods will be used for projecting benefits and different discount rates will be employed for estimating present values. There will be similar differences in the techniques for cost measurements. In order to make the results of the studies comparable, all of the assumptions and techniques need to be equated. It is our hope that if evaluators follow the ideas and procedures presented in this primer, there will be sufficient similarity among individual studies to make them comparable.[21]

---

[21] The best alternative appears to be to conduct multiprogram evaluations where a conscious effort is made to evaluate more than one program using the same techniques, assumptions, time period, and so forth. While methodologically this procedure is preferred, the simultaneous evaluation of a number of programs requires considerable resources, and administering such a project becomes a major job which may flounder due to its complexities.

A Final Word for Impact Evaluation

The type of analysis we have proposed is difficult to conduct and obviously filled with pitfalls. It may not be correct when completed. There are reasons for its use, however. It forces those responsible for social program decisions to attach weights to their goals and to quantify the success and costs of a program as far as is possible rather than rest content with vague qualitative judgments and personal hunches. This is obviously a good thing in itself; some information is better than none. Also, it has the very valuable by-product of raising questions which would otherwise not have been asked. There is a considerable expansion in the outlook with which the programs are viewed, from daily operational questions to the broader perspectives of social impacts and costs. Thus, even though impact evaluation may not always give the "right" answers, it may lead to the asking of more "right" questions if used sensibly. As experience and expertise are accumulated, this method should lead to better answers.

Exercise 5-1

Amy R. has recently completed high school and is contemplating her future. She has several alternatives from which to choose.

1) She has been offered a job as a sales trainee by a large manufacturing firm. They will start her at a salary of $7,000 per year with guaranteed increases of $1,000 for the next two years. After that point she will be paid commissions only. Successful salespersons earn $25,000 per year, but only 10 percent of all trainees will eventually fit into this classification. An additional 40 percent average $15,000 per year. The remainder average $10,000 per year.

2) She can enter an apprenticeship to become a carpenter. Under the terms of the apprenticeship she will earn $3.50, $4.25, and $5.00 per hour, respectively, for the three years of the apprenticeship and expect to work about 1,800 hours per year. When she completes the program and receives her union card, she will make $7.00 per hour for 2,000 hours of work.

3) She can enter college and study to be an accountant. The four years of school will cost $3,000 per year in tuition, room, board, and other out-of-pocket costs. Accountants earn about $12,000 to start and increase to average earnings of $15,000, $20,000, and $30,000 per year after five, ten, and twenty years of experience, respectively.

Ignoring her abilities and interest, what considerations should Amy R. consider in choosing among the alternatives? Under what circumstances would each of the alternatives be preferable?

Exercise 5-2

The National Institute for Conserving Energy (NICE) has awarded the City of Podunk a $1,000,000 grant to hire unemployed older persons to winterize and insulate homes and apartments in low income neighborhoods of Podunk. The workers are paid the minimum wage and the materials are purchased directly from the national manufacturers at the manufacturers' cost.

Congressman Joe Nathan who represents the district which includes Podunk has received the following letter signed jointly by the Podunk Insulating Contractors Association, Local 57 of the United Insulators Union, and the Building Suppliers Association, criticizing the program:

Dear Congressman Nathan:

As a champion of the free enterprise system, we know you will want to deal with yet another example of federal interference in the free market. Several months ago your office announced a large grant to the City of Podunk which was designed to reduce unemployment from its present high rate of 12 percent by hiring the unemployed to carry on insulating work.

Mr. Congressman, this project is creating more unemployment than it is solving. We insulate homes throughout the city, in all of its neighborhoods. By the hiring of the unemployed older workers at substandard wages, the government is substituting their labor for those of our members. Furthermore, by purchasing supplies from the manufacturers, the city is depriving local building suppliers from providing their normal services.

The result of these actions could be ruinous to the insulating industry of Podunk. Mr. Nathan, we are counting on you to stop this bureaucratic blundering and save our jobs.

Congressman Nathan has forwarded the letter to the Director of NICE. You have been instructed to examine the issue and write a reply for the Director.

1) What additional information would you like to have about the situation?
2) Write the reply assuming that you had all of the relevant facts.

REFERENCES

1. Ashenfelter, Orley. "The Effect of Manpower Training on Earnings: Preliminary Results," Technical Analysis Paper No. 12B, Office of Evaluation, Office of the Assistant Secretary for Policy, Evaluation and Research, June 1975.

2. Ashenfelter, Orley. "Estimating the Effects of Training Programs on Earnings," *The Review of Economics and Statistics.* Vol. 60, No. 1 (February 1978), pp. 47-57.

3. Boruch, Robert F. "On Common Contentions About Randomized Field Experiments," *Evaluation Studies Review Annual.* Gene V. Blass, ed. Vol. 1. Beverly Hills: Sage Publications, 1976, pp. 158-194.

4. Boruch, Robert F. and Henry W. Riecken, eds. *Social Experimentation: A Method for Planning and Evaluating Social Intervention.* New York: Academic Press, 1975.

5. Borus, Michael E. "Time Trends in the Benefits from Retraining in Connecticut," *Proceedings of the Twentieth Annual Winter Meeting.* Madison, WI: Industrial Relations Research Association, 1968, pp. 36-46.

6. Borus, Michael E., ed. *Evaluating the Impact of Manpower Programs.* Lexington, MA: D. C. Heath and Co., 1972.

7. Borus, Michael E. and Edward C. Prescott. "The Effectiveness of MDTA Institutional Training Over Time and in Periods of High Unemployment," *American Statistical Association 1973 Proceedings of the Business and Economic Statistics Section.* Washington, DC: American Statistical Association, pp. 278-284.

8. Cain, Glen G. "Regression and Selection Models to Improve Nonexperimental Claims," *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs.* Carl A. Bennett and Arthur A. Lumsdaine, eds. New York: Academic Press, 1975.

9. Campbell, Donald T. "Reforms as Experiments," *Handbook of Evaluation Research.* Elmer Struening and Marcia Guttentag, eds. Vol. 1. Beverly Hills: Sage Publications, 1975, pp. 71-100.

10. Campbell, Donald T. and Robert F. Boruch. "Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effect," *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs.* Carl A. Bennett and Arthur A. Lumsdaine, eds. New York: Academic Press, 1975.

11. Campbell, Donald T. and Albert Erlebacher. "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful," *Handbook of Evaluation Research.* Elmer Struening and Marcia Guttentag, eds. Vol. 1. Beverly Hills: Sage Publications, 1975, pp. 597-617.

12. Campbell, Donald T. and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally & Co., 1971.

13. Director, Steven M. "Underadjustment Bias in the Evaluation of Manpower Training," a report submitted to the Office of Evaluation, ASPER, U.S. Department of Labor, NTIS Accession Number PB 270525, Springfield, VA: National Technical Information Service, 1976.

14. Donnelly, Bernard P. "Cheap Shots and Costly Pay-Offs: A Plea for Purpose in Public Programs," *Public Administration Review.* No. 2 (March/April 1977), pp. 181-186.

15. Ferman, Louis A. "Some Perspectives on Evaluating Social Welfare Programs," *The Annals of the American Academy of Political and Social Science.* Vol. 385 (September 1969), pp. 143-156.

16. *Focus on Manpower Planning: 2-Monitoring and Evaluation.* Washington, DC: National League of Cities, United States Conference of Mayors, December 1973.

17. Hardin, Einar and Michael E. Borus. *The Economic Benefits and Costs of Retraining.* Lexington, MA: D. C. Heath and Co., 1971.

18. Heckman, James J. "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement.* Vol. 5, No. 4 (Fall 1976), pp. 475-492.

19. Hu, Teh-wei, Maw Lin Lee, and Ernst W. Stromsdorfer. "Economic Returns to Vocational and Comprehensive High School Graduates," *The Journal of Human Resources.* Vol. VI, No. 1 (Winter 1971), pp. 25-50.

20. Judy, Richard W. "Costs: Theoretical and Methodological Issues," *Cost-Benefit Analysis of Manpower Policies.* G. G. Somers and W. D. Wood, eds. Kingston, Ontario: Industrial Relations Centre, Queen's University, 1969, pp. 16-29.

21. Menzi, Donald. "Developing a Performance Rating System for Manpower Training Programs," *Evaluation.* 2(2), 1975, pp. 50-54.

22. Pitcher, Hugh M. "A Sensitivity Analysis to Determine Sample Sizes for Performing Impact Evaluation of the CETA Programs," *Evaluating Manpower Training Programs.* Farrell Bloch, ed. Greenwich, CT: JAI Press, 1978.

23. Prest, A. R. and Ralph Turvey. "Cost-Benefit Analysis: A Survey," *Surveys of Economic Theory.* Vol. III. New York: St. Martin's Press, 1967.

24. Robinson, John P., Robert Athanasiou, and Kendra B. Head. *Measures of Occupational Attitudes and Occupational Characteristics.* Ann Arbor: Institute for Social Research, University of Michigan, 1973.

25. Robinson, John P. and Phillip R. Shaver. *Measures of Social Psychological Attitudes*. Ann Arbor: Institute for Social Research, University of Michigan, 1973.

26. Robinson, John P., Jerrold G. Rusk, and Kendra B. Head. *Measures of Political Attitudes*. Ann Arbor: Institute for Social Research, University of Michigan, 1969.

27. Salasin, Susan. "Exploring Goal-Free Evaluation: An Interview with Michael Scriven," *Evaluation*. 2(1), 1974, pp. 9-16.

28. Scriven, Michael. "The Methodology of Evaluation," *Perspectives of Curriculum Evaluation: AERA Monograph Series on Curriculum Evaluation*. Chicago: Rand McNally, 1967.

29. Scriven, Michael. "Pros and Cons About Goal-Free Evaluation," *The Journal of Educational Evaluation*. Vol. 3, No. 4 (December 1972), pp. 1-7.

30. Smith, Ralph E. "The Opportunity Cost of Participating in a Training Program," *Journal of Human Resources*. Vol. VI, No. 4 (Fall 1971), pp. 510-519.

31. Somers, Gerald G. and Graeme H. McKechnie. "Vocational Retraining Programs for the Unemployed," *Proceedings of the Twentieth Annual Winter Meeting*. Madison, WI: Industrial Relations Research Association, 1968, pp. 25-35.

32. Somers, Gerald G. and Ernst W. Stromsdorfer. *A Cost-Effectiveness Study of the In-School and Summer Neighborhood Youth Corps*. Madison, WI: Industrial Relations Research Institute, University of Wisconsin, 1970.

33. Suits, Daniel B. *Statistics: An Introduction to Quantitative Economic Research*. Chicago: Rand McNally & Co., 1971.

34. U.S. Department of Labor, Employment and Training Administration. *Self Evaluation of CETA Manpower Programs: A Guide to Prime Sponsors*. Springfield, VA: National Technical Information Service, 1976.

35. Weiss, Carol H. "Interviewing in Evaluation Research," *Handbook of Evaluation Research.* Elmer L. Struening and Marcia Guttentag, eds. Vol. 1. Beverly Hills: Sage Publications, 1975, pp. 355-395.

36. Wholey, Joseph S., John W. Scanlon, Hugh G. Duffy, James S. Fukumoto, and Leona M. Vogt. *Federal Evaluation Policy: Analyzing the Effects of Public Programs.* Washington, DC: The Urban Institute, 1970.